

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
2 September 2004 (02.09.2004)

PCT

(10) International Publication Number
WO 2004/074429 A2

(51) International Patent Classification⁷: C12N

(21) International Application Number:
PCT/DK2004/000117

(22) International Filing Date: 23 February 2004 (23.02.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

PA 2003 00268	21 February 2003 (21.02.2003)	DK
PA 2003 00269	21 February 2003 (21.02.2003)	DK
60/448,480	21 February 2003 (21.02.2003)	US
60/448,460	21 February 2003 (21.02.2003)	US
PA 2003 01356	18 September 2003 (18.09.2003)	DK
60/504,748	22 September 2003 (22.09.2003)	US

(71) Applicant (for all designated States except US): NUEVO-LUTION A/S [DK/DK]; Rønnegade 8,5. sal, DK-2100 København Ø (DK).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **FRESKGÅRD, Per-Ola** [SE/SE]; Örtugsgatan 50, S-603 79 Norrköping (SE). **GOULIAEV, Alex, Haahr** [DK/DK]; Brødsted 223, DK-3670 Veksø Sjælland (DK). **THISTED, Thomas** [DK/DK]; Fjordskrænten 14, DK-3600 Frederikssund (DK). **OLSEN, Eva, Kampmann** [DK/DK]; Vingetoften 17, DK-2730 Herlev (DK).

(74) Agent: **HØIBERG A/S**; Store Kongensgade 59 A, DK-1264 Copenhagen K. (DK).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI,

GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:

— as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHOD FOR PRODUCING SECOND-GENERATION LIBRARY

(57) Abstract: The present invention relates to a method for generating a second-generation library. In a first step, a library of encoded molecules associated with an identifier nucleic acid comprising codons identifying chemical entities that have participated in the formation of the encoded molecule is provided. In a second step, the library is partitioned and encoded molecules having a certain property are selected. Codons of identifiers of selected encoded molecules are subsequently identified, and a second-generation library is prepared using at least some of the chemical entities coded for by the identified codons. The new focussed library may be used for another partition step to select encoded molecules with a certain property.

Title**METHOD FOR PRODUCING SECOND-GENERATION LIBRARY**

Various patent and non-patent references cited in the present application are hereby
5 incorporated by reference in their entirety.

Technical Field of the Invention

The invention relates to a method for producing a second-generation compound
library with an improved desired property profile. In nature and artificial methods
10 based on the natural system, the parent genotype is carried on to the off-spring and
results in a phenotype in which the exact type and sequence of amino acids is
retained, unless a mutation and/or recombination has occurred. The present method
only retains the identity of chemical entities, e.g. amino acids, while the sequence
wholly or partly is scrambled. The result is a focused second-generation library with
15 lower diversity.

Background of the Invention

The biological evolution is based on the survival of specific genotypes that encode
phenotypes with the most suitable functionalities in a certain environment. In all liv-
20 ing species DNA programs the genotype. DNA serves two important functions in the
natural selection process. One function is obviously to encode for the type of nucleo-
tides used and the other function is to encode for the specific order of nucleotide
sequences in a nucleic acid sequence. The strategy used in nature, i.e. encoding for
the exact type as well as the precise sequence of nucleotides, ensures an extremely
25 similarity between the progeny and its parents. Thus, conserving almost the exact
sequence and type of the nucleotides is absolutely essential in order to create off
spring with a high functionality. The changes in the genotype from one generation
to another, which allow for evolution, are determined by the random mutation rate
and recombination between the two parent's genotypes.

30

The natural selection cannot afford too many changes in the DNA from one genera-
tion to the next in order to secure survival of the species. Therefore, nature has
evolved sophisticated means to proofread the copying of the DNA from the parents
to its progeny and secured that the characteristics of phenotype from one generation
35 to the next is carried only by the DNA.

Within the art of selecting ligands from a library of encoded polypeptides associated with a corresponding identifier nucleic acid sequence, the method of nature is used. Thus, when more than a single library generation is needed, the identifier nucleic acid sequences (genotype) carries the information from one generation to the next.

WO 93/03172 A1 discloses a method for identifying a polypeptide ligand having a desired property in a polypeptide library. In a first step, a translatable mRNA mixture is provided, which is mixed with a mixture of ribosome complexes to form a translation product attached to the mRNA strand responsible for the formation thereof. In a second step the ribosome complexes binding to a target are partitioned from and remainder of the library. In a third step, an amplification of mRNA strands of the partitioned ribosome complexes, which has bound to the target follows. The amplified mRNA strands are used for the production of a second generation library, which is subjected to a renewed contact with the target. The method is repeated a sufficient number of times until the size of the library has narrowed to a small pool of high affinity binders.

In WO98/31700 A1 a method for selecting a DNA molecule, which encodes for a desired protein, is disclosed. The method implies the initial presence of a pool of candidates RNA molecules, which subsequently is translated into a corresponding pool of RNA-protein fusions. Subsequently the mRNA-protein fusion products are subjected to a selection process, i.e. the fusion products are presented for a target molecule, and a new pool of complexes capable of binding to the target are partitioned. From the new pool of complexes, the mRNAs are recovered and amplified for use in a subsequent round of library generation. Xu, L. et al *Chemistry & Biology*, Vol. 9, 933-942, August 2002 discloses a practical embodiment in which a library of more than 10^{12} unique mRNA-protein fusion products through ten rounds of library generation and selection are used to identify a high affinity binding protein.

The preparation of libraries of synthetic molecules associated with a corresponding identifier nucleic acid sequence, and the selection of synthetic molecules from such libraries, have been the subject of various patent applications. When two or more generations of libraries are needed, the identifier nucleic acid sequence is used as carrier between an initial library and the next generation library.

Thus, in WO 00/23458A1 libraries of complexes comprising non-natural molecules attached to corresponding nucleic acid sequences are suggested. After a selection of the library has been conducted, the nucleic acid sequences of successful complexes are amplified by PCR and a new library is prepared from these nucleic acid sequences. The same method of carrying information from an initial library to the next library is applied in WO 02/074929A2 and WO 02/103008A2.

The present invention provides a new method for evolving encoded molecules. The method is based on the identification of chemical entities used in the synthesis of reaction products of successful complexes and the application, at least in part, of these chemical entities in the preparation of the next generation library. The utilization of preferable chemical entities and the exclusion of certain undesired chemical entities in the next library generation generally imply that the next generation library has a smaller size compared to the size of the initial library, thereby, at the same time, retaining the desirable encoded molecules in the library.

Summary of the Invention

The present invention concerns a method for producing a composition of molecules with an improved desired property, said method comprising the steps of: providing an initial library comprising a plurality of different encoded molecules associated with a corresponding identifier nucleic acid sequence, wherein each encoded molecule comprises a reaction product of multiple chemical entities and the identifier nucleic acid sequence comprises codons identifying said chemical entities; subjecting the initial library to a condition partitioning members having encoded molecules displaying a predetermined property from the remainder of the initial library; identifying codons of the identifier nucleic acid sequences of the partitioned members of the initial library; and preparing a second-generation library of encoded molecules using the chemical entities coded for by the codons of the partitioned members of the initial library or a part thereof.

The present invention relates to a novel approach to perform evolution of molecules with a desired property, said approach being different from the approach of nature and the prior art. The invention is based on the selecting of chemical entities, the counterpart of amino acids in Nature, instead of the precise sequence of chemical

entities. This new approach is powerful in *ex vivo* conditions when high functionality of the off spring is not vital for success and when the number of chemical entities relative to the number of reactants used in each encoded molecule is high.

5 The method disclosed herein will be increasingly effective as the library size increases. This is due to the fact that more chemical entities is used when a library size is increased, when the number of reactions for the formation of the encoded is fixed and the fact that different chemical entities tend to be involved in encoded molecules having the desired property. The chemical entities, which are part of the
10 final selected molecules, will be enriched in each round of selection. Finally, when the diversity has been extensively reduced, the enriched molecules are decoded from the identifier nucleic acid sequence comprising the codons of the chemical entities that have participated in the formation of the encoded molecule.

15 The strategy of performing enrichment of chemical entities instead of specific combinations of chemical entities more efficiently search the chemical space for all combinations of chemical entities that are eager to show a certain property, such as a binding ability towards a target. Thus, chemical entities having a certain impact on the formation of encoded molecules is allowed in a new library to recombine in each
20 new library generation. In a certain aspect of the invention, the recombination is random, i.e. once a chemical entity has qualified as being of interest it is allowed in every position of the reaction sequence. In another aspect of the invention, the recombination is semi-random, i.e. once a chemical entity is qualified as being of interest it is used in a certain position in the reaction sequence of the encoded molecule. In still a further aspect of the invention, the amount of the chemical entity used
25 in a subsequent library generation is dependent on the frequency and the amount of the partitioned library members.

The present invention may be of special interest when a group of chemical entities
30 are selected from a larger pool of chemical entities in the formation of a first library. Selecting chemical entities resulting in encoded molecules having a certain property in a first library and spiking with remaining chemical entities of the pool allows for the formation of a second-generation library not necessarily of a smaller size but enriched in encoded molecules having a certain property.

The second-generation library may be formed of a reaction product of the chemical entities without attaching the reaction product to a nucleic acid. In an embodiment of such second-generation library the individual reaction products are formed in discrete reaction compartments in accordance with traditional combi-chem technology.

5 In a certain aspect of the invention, the second-generation library is prepared as the first generation library, i.e. the second-generation library comprises a plurality of different encoded molecules associated with a corresponding identifier nucleic acid sequence, wherein each encoded molecule comprises a reaction product of multiple chemical entities and the identifier nucleic acid sequence comprises codons identifying said chemical entities.

10 In a preferred aspect of the invention, it comprises subjecting the second-generation library to a condition partitioning members having encoded molecules displaying a predetermined property from the remainder of the second-generation library. The second-generation library may be partitioned as to the same property or a different property. Notably, the second-generation library can be screening against the same target or a different target.

20 After the partitioning of the second-generation library, the invention comprises the step of deducing the identity of the encoded molecule(s) using the identifier nucleic acid sequence, when present. Optionally, a third or further generation library may be formed and screened before the final deducing step is performed. In a certain embodiment, the decoding includes that the codons of the identifier nucleic acid sequence is decoded to establish the synthesis history of the encoded molecules. The synthesis history includes the identity of the chemical entities used and the point in time they enter the sequence of reactions resulting in the encoded molecule.

30 The encoded molecule is preferably a reaction product in which multiple chemical entity precursors have participated. The encoded molecule may have any chemical structure. Generally, the multiple chemical entities are precursors for a structural unit appearing in the encoded molecule. However, the chemical entities may also perform a chemical reaction with the nascent encoded molecule, which result in an altering or removal of chemical groups. In certain aspects of the invention, the encoded molecule is a scaffolded molecule, i.e. various chemical entities have reacted with a chemical core structure like steroid, benzodiazepine, retinol,

camphor, ephedrine, penicillin, cannabinol, coumarin, oxazol, etc. In certain other aspects of the invention the encoded molecule is fully or partly a polymer. The polymer may be of a type which occurs naturally or may be a non-naturally occurring polymer. Nature only has the possibility of preparing α -polypeptides using the
5 recognition of a codon of an mRNA strand by the anticodon of a charged tRNA. In some aspects of the invention, the encoded molecule is not a α -peptide. Notably, in some aspects of the invention, the chemical entities are reacted without enzymatic interaction to produce the encoded molecule.

10

The encoded molecule can be associated with the nucleic acid sequence identifier in any appropriate way. In a certain aspect of the invention, the encoded molecule associated with the corresponding identifier nucleic acid sequence is a bifunctional complex. The bifunctional complex may be formed by covalent or non-covalent at-
15 tachment of the encoded molecule to the identifier nucleic acid sequence. In another aspect of the invention, an identifier nucleic acid sequence is physically a distinct entity separated from the encoded molecule, wherein the identifier identifies the spatial position of an encoded molecule, e.g. in the same compartment in which an encoded molecule is formed a corresponding identifier oligonucleotide is generated.

20

The conditions partitioning complexes of interests from the remainder of the library may be chosen from a variety of possibilities. In one aspect the condition relates to physical parameters, so that complexes displaying a physical stability under e.g. certain temperature conditions, certain acidic conditions, certain radiation conditions
25 etc. are selected from the library. In other aspects of the invention the condition for partitioning the desired complexes includes subjecting the initial library to a molecular target and partitioning complexes binding to this target. The molecular target may be any compound of interest. Exemplary targets are proteins, carbohydrates, polysaccharides, hormones, receptors, antibodies, viruses, antigens, cells, tissues etc.

30

In certain aspects the target is immobilized on a solid support, such as column material and contacted with the candidate complexes in a fluid media followed by a partitioning of the complexes capable of binding to the target under the contacting conditions used. Typically the binding complexes are eluted from the column using increased stringency conditions.

35

The complexes as such or only the identifier part is harvested after the partitioning step. Usually the identifier nucleic acid sequences are amplified prior to the identification step. The amplification is suitably performed applying polymerase chain reaction (PCR). The amplified identifiers may be explicitly or implicitly identified. When
5 the codons are identified explicitly, the sequence and identity of nucleotides in the codon is made known to the experimenter, whereas, when the codons of the identifiers are implicitly identified, the experimenter is not presented for the information.

Any suitable method for identifying codons may be used. In a certain aspect of the
10 invention, traditional sequencing, e.g. by using a modification of the Sangers method or pyrosequencing methods, identifies the codons. In another aspect of the invention, the codons of the identifier nucleic acid sequences of the partitioned members of the initial library are identified by contacting said identifier nucleic acid sequences with a pool of nucleic acid fragments under conditions allowing for hybridisation.

15 The pool of nucleic acid fragments may be immobilized or in solution. In a certain aspect of the invention, the pool of nucleic acid fragments comprises a plurality of single stranded nucleic acid probes immobilized in discrete areas of a solid support, wherein the nucleic acid probes are capable of hybridising to a codon of the
20 identifier nucleic acid sequence comprising codons. The nucleic acid probes may be positioned on a microarray, such that the identity of the codons is revealed by observing the discrete areas of the support in which a hybridisation event has occurred.

25 The nucleic acid probe can be directly hybridised to the identifier or the nucleic acid probe of the array is hybridised to an identifier nucleic acid sequence through an adapter oligonucleotide having a sequence complementing the probe as well as one or more codons of the identifier nucleic acid sequence. The probe may identify a single codon of an identifier or a probe of the array is capable of hybridising to two
30 codons of the identifier nucleic acid sequence or a sequence complementary to said sequence. The ability to hybridise two or more codons makes it possible to study the influences of neighbouring chemical entities on each other. In a certain aspect, a nucleic acid probe of the array is capable of hybridising to all codons of an identifier nucleic acid sequence. This latter option will fully decode the identity of the encoded

molecule. Usually however, a fully decoding is only possible for a relative small library size, as it presupposes a nucleic acid probe for each member of the library.

When single codons are detected, useful information about a certain codon may be gathered by detecting the codon together with a framing sequence identifying the position in the reaction history of the chemical entity corresponding to said codon.

As an example, if a library of complexes is prepared from 100 chemical entities and the three reactions, i.e. each identifier comprises 4 codons, the library size is 10^8 . For most practical uses 10^8 is in the excess of what is possible to detect on an array, especially if multiple determinations for each identifier are considered necessary to obtain a high accuracy. However, an array of just 100 probes complementary to the 100 codons will reveal important information prior to or subsequent to a selection. In the event a framing sequence is detected together with the codon an array of 400 probes is needed.

A suitable method for identifying an hybridisation event is to use a label. Therefore, in a preferred embodiment, the existence of a hybridisation event is measured through labelling of the identifier nucleic acid sequence, or an amplification product thereof. When the label emits light, the hybridisation event is measured by the emission of light in a scanner. To reveal the relative abundance of each chemical entity in the library of encoded molecules, the relative intensity of light in each discrete spot is measured.

The measurement of a hybridisation event may be conducted by various methods known in the art. In the event the label emits lights, the presence or absence of a hybridisation event may be measured in a scanner, e.g. a confocal scanner. The scanner may be connected with computer software, which is able to quantify the amount of lights measured. The amount of light measured correlates with the amount of identifier annealed to the probes. Thus, it is possible to measure not only the presence or absence of one or more codons of an identifier ; it is also possible to measure the relative amount of the codons in one or more identifiers.

After the complexes have been partitioned and the specific codons have been identified on the microarray, the information can be used to design optimized libraries

including chemical entities based on both the selection data and the chemical structure. The microarray analysis will first of all detect which chemical entities pass the partitioning step. Secondly, the relative intensity on the microarray will reflect the relative binding affinity of the chemical entities. Finally, the structures of the chemical entities are directly identified due to the position of the probes on the array. For instance, chemical entities that are strongly selected in a partitioning process but possess some unfavourable chemical structure can be excluded in the next generation of library. Similarly, chemical entities that are weakly selected in a partitioning process but possess some favourable chemical structure can be included in the next generation of library. Thus, the next generation library design can be based both on a rational choice of chemical entities with lead-like structures and the selection pressure detected on the microarray.

Another method of identifying codons includes that nucleic acid fragments are primer oligonucleotides, and the identification involves subjecting the hybridisation complex between the primer oligonucleotides and the identifier nucleic acid sequences to a condition allowing for an extension reaction to occur when the primer is sufficient complementary to a part of the identifier nucleic acid sequence, and evaluating based on measurement of the extension reaction, the presence, absence, or relative abundance of one or more codons.

The extension reaction requires a primer, a polymerase as well as a collection of deoxyribonucleotide triphosphates (abbreviated dNTP's herein) to proceed. An extension product may be obtained in the event the primer is sufficient complementary to an identifier oligonucleotide for a polymerase to recognise the double helix as a substrate. After binding of the polymerase to the double helix, the deoxyribonucleotide triphosphates (blend of dATP, dCTP, dGTP, and dTTP) are incorporated into the extension product using the identifier oligonucleotide as identifier. The conditions allowing for the extension reaction to occur usually includes a suitable buffer. The buffer may be any aqueous or organic solvent or mixture of solvents in which the polymerase has a sufficient activity. To facilitate the extension process the polymerase and the mixture of dNTP's are generally included in a buffer which is added to the identifier oligonucleotide and primer mixture. An exemplary kit comprising the polymerase and the nNTP's for performing the extension process comprises the following: 50 mM KCl; 10 mM Tris-HCl at pH 8.3; 1.5 mM MgCl₂ ; 0.001% (wt/vol)

gelatin, 200 μ M dATP; 200 μ M dTTP; 200 μ M dCTP; 200 μ M dGTP; and 2.5 units *Thermus aquaticus* (Taq) DNA polymerase I (U.S. Pat. No. 4,889,818) per 100 microliters (μ l) of buffer.

- 5 The primer may be selected to be complementary to one or more codons or parts of such codons. The length of the primers may be determined by the length of the codons, however, the primers usually are at least about 11 nucleotides in length, more preferred at least 15 nucleotides in length to allow for an efficient extension by the polymerase. The presence or absence of one or more codons is indicated by the
10 presence of or absence of an extension product. The extension product may be measured by any suitable method, such as size fractioning on an agarose gel and staining with ethidium bromide.

- In a preferred embodiment the admixture of identifier oligonucleotide and primer is
15 thermocycled to obtain a sufficient number of copies of the extension product. The thermocycling is typically carried out by repeatedly increasing and decreasing the temperature of the mixture within a temperature range whose lower limit is about 30 degrees Celsius (30°C) to about 55°C and whose upper limit is about 90°C to about 100° C. The increasing and decreasing can be continuous, but is preferably phasic
20 with time periods of relative temperature stability at each of temperatures favouring polynucleotide synthesis, denaturation and hybridization.

- When a single complex is analysed in accordance with the present method, the result may be used to verify the presence or absence of a specific chemical entity during the formation of the display molecule. The formation of an extension product is
25 indicative of the presence of an oligonucleotide part complementary to the primer in the identifier oligonucleotide. Conversely, the absence of an extension product is indicative of the absence of an oligonucleotide part complementary to the primer in the identifier oligonucleotide. Selecting the sequence of the primer such that it is
30 complementary to one or more codons will therefore provide information of the structure of the encoded molecule coded for by this codon(s).

- In a preferred aspect of the invention, in the mixture of the identifier oligonucleotide and the primer oligonucleotide, a second primer complementary to a sequence of
35 the extension product is included. The second primer is also termed reverse primer

and ensures an exponential increase of the number of produced extension products. The method using a forward and reverse primer is well known to skilled person in the art and is generally referred to as polymerase chain reaction (abbreviated PCR) in the present application with claims. In one embodiment of the invention the reverse primer is annealed to a part of the extension product downstream, i.e. near the 3' end of the extension product, or a part complementing the coding part of the identifier oligonucleotide. In another embodiment, the first primer (forward primer) anneals to an upstream position of the identifier oligonucleotide, preferably before the coding part, and the reverse primer anneals to a sequence of the extension product complementing one or more codons or parts thereof.

The amplicons resulting from the PCR process may be stained during or following the reaction to ease the detection. A staining after the PCR process may be prepared with e.g. ethidium bromide or a similar staining agent. As an example, amplicons from the PCR process is run on an agarose gel and subsequently stained with ethidium bromide. Under UV illumination bands of amplicons becomes visible. It is possible to incorporate the staining agent in the agarose gel or to allow a solution of the staining agent to migrate through the gel. The amplicons may also be stained during the PCR process by an intercalating agent, like CYBR. In presence of the intercalating agent while the amplification proceeds it will incorporate in the double helix. The intercalation agent may then be made visible by irradiation by a suitable source.

The intensity of the staining is informative of the relative abundance of a specific amplicon. Thus, it is possible to quantify the occurrence of a codon in an identifier oligonucleotide. When a library of bifunctional complexes has been subjected to a selection the codons in the pool of identifier oligonucleotides which has been selected can be quantified using this method. As an example a sample of the selected identifier oligonucleotides is subjected to various PCR amplifications with different primers in separate compartments and the PCR product of each compartment is analysed by electrophoresis in the presence of ethidium bromide. The bands that appear can be quantified by a densitometric analysis after irradiation by ultraviolet light and the relative abundance of the codons can be measured.

Alternatively, the primers may be labelled with a suitable small molecule, like biotin or digoxigenin. A PCR-ELISA analysis may subsequently be performed based on the amplicons comprising the small molecule. A preferred method includes the application of a solid support covered with streptavidin or avidin when biotin is used as
5 label and anti-digoxigenin when digoxigenin is used as the label. Once captured, the amplicons can be detected using an enzyme-labelled avidin or anti-digoxigenin reporter molecule similar to a standard ELISA format.

To avoid laborious post-PCR handling steps required to evaluate the amplicons, it is
10 in a certain embodiment preferred to measure the extension process "real time". Several real time PCR processes have been developed and all the suitable real time PCR process available to the skilled person in the art can be used in the evaluating step of the present invention and are included in the present scope of protection. The PCR reactions discussed below are of particular interest.

15 The monitoring of accumulating amplicons in real time has been made possible by labelling of primers, probes, or amplicons with fluorogenic molecules. The real time PCR amplification is usually performed with a speed faster than the conventional PCR, mainly due to reduced cycles time and the use of sensitive methods for detection of emissions from the fluorogenic labels. The most commonly used fluorogenic
20 oligoprobes rely upon fluorescent resonance energy transfer (FRET) between fluorogenic labels or between one fluorophore and a dark or "black-hole" non-fluorescent quencher (NFQ), which disperse energy as heat rather than fluorescence. FRET is a spectroscopic process by which energy is passed between molecules separated by 10-100 Å that have overlapping emission and absorption spectra. An advantage of many real time PCR methods is that they can be carried out in
25 a closed system, i.e. a system which does not need to be opened to examine the result of the PCR. A closed system implies a reduced result turnaround, minimisation of the potential for carry-over contamination and the ability to closely scrutinise the assay's performance.
30

The real time PCR methods currently available to the skilled person can be classified into either amplicon sequence specific or non-specific methods. The basis for the non-specific detection methods is a DNA-binding fluorogenic molecule. Included
35 in this class are the earliest and simplest approaches to real time PCR. Ethidium

bromide, YO-PRO-1, and SYBR[®] green 1 all fluorescence when associated with double stranded DNA which is exposed to a suitable wavelength of light. This approach requires the fluorescent agent to be present during the PCR process and provides for a real time detection of the fluorescent agent as it is incorporated into the double stranded helix.

The amplicons sequence specific methods includes, but are not limited to, the TaqMan[®], hairpin, LightCycler[®], Sunrise[®], and Scorpion[®] methods. The LightCycler[®] method also designated "HybProbes" make use of a pair of adjacent, fluorogenic hybridisation oligonucleotide probes. A first, usually the upstream oligoprobe is labelled with a 3' donor fluorophore and the second, usually the downstream probe is commonly labelled with either a Light cycler Red 640 or Red 705 acceptor fluorophore at the 5' terminus so that when both oligoprobes are hybridised the two fluorophores are located in close proximity, such as within 10 nm, of each other. The close proximity provides for the emission of a fluorescence when irradiated with a suitable light source, such a blue diode in case of the LightCycler[®]. The region for annealing of the probes may be any suitable position that does not interfere with the primer annealing. In a suitable setup, the site for binding the probes are positioned downstream of the codon region on the identifier oligonucleotide. Alternatively, when a reverse primer is used, the region for annealing the probes may be at the 3' end of the strand complementing the identifier oligonucleotide. Another embodiment of the LightCycler method includes that the pair of oligonucleotide probes are annealed to one or more codons and primer sites exterior to the coding part of the identifier oligonucleotide are used for PCR amplification.

The TaqMan[®] method, also referred to as the 5' nuclease or hydrolysis method, requires an oligoprobe, which is attached to a reporter fluorophore, such as 6-carboxy-fluorescein, and a quencher fluorophore, such as 6-carboxy-tetramethyl-rhodamine, at each end. When in close proximity, i.e. annealed to an identifier oligonucleotide, or a sequence complementing the identifier oligonucleotide, the quencher will "hijack" the emissions that have resulted from the excitation of the reporter. As the polymerase progresses along the relevant strand, it displaces and hydrolyses the oligoprobe via its 5'→3' endonuclease activity. Once the reporter is removed from the extinguishing influence of the quencher, it is able to release excitation energy at a wavelength that can be monitored by a suitable instrument,

such as ABI Prism® 7700. The fractional cycle number at which the real-time fluorescence signal mirrors progression of the reaction above the background noise is normally used as an indicator of successful identifier oligonucleotide amplification. This threshold cycle (C_T) is defined as the PCR cycle in which the gain in fluorescence generated by the accumulating amplicons exceeds 10 standard deviations of the mean base line fluorescence. The C_T is proportional to the number of identifier oligonucleotide copies present in the sample. The TaqMan probe is usually designed to hybridise at a position downstream of a primer binding site, be it a forward or a reverse primer. When the primer is designed to anneal to one or more codons of the identifier oligonucleotide, the presence of these one or more codons is indicated by the emittance of light. Furthermore, the quantity of the identifier oligonucleotides comprising the one or more codons may be measured by the C_T value.

The Hairpin method involves an oligoprobe, in which a fluorophore and a quencher are positioned at the termini. The labels are held in close proximity by distal stem regions of homologous base pairing deliberately designed to create a hairpin structure which result in quenching either by FRET or a direct energy transfer by a collisional mechanism due to the intimate proximity of the labels. When direct energy transfer by a collision mechanism is used the quencher is usually different from the FRET mechanism, and is suitably 4-(4'-dimethylamino-phenylazo)-benzene (DAB-CYL). In the presence of a complementary sequence, usually downstream of a primer, or within the bounds of the primer binding sites in case of more than one a single primer, the oligoprobe will hybridise, shifting into an open configuration. The fluorophore is now spatially removed from the quencher's influence and fluorescence emissions are monitored during each cycle. In a certain aspect, the hairpin probe may be designed to anneal to a codon in order to detect this codon if present on the identifier oligonucleotide. This embodiment may be suitable if codons only differs from each other with a single or a few nucleotides, because it is well-known that the occurrence of a mismatch between a hairpin oligoprobe and its target sequence has a greater destabilising effect on the duplex than the introduction of an equivalent mismatch between the target oligonucleotide and a linear oligoprobe. This is probably because the hairpin structure provides a highly stable alternate conformation.

The Sunrise and Scorpion methods are similar in concept to the hairpin oligoprobe, except that the label becomes irreversible incorporated in to the PCR product. The Sunrise method involves a primer (commercially available as Amplifluor™ hairpin primers) comprising a 5' fluorophore and a quencher, e.g. DABCYL. The labels are
5 separated by complementary stretches of sequence that create a stem when the sunrise primer is closed. At the 3' terminus is a target specific primer sequence. In a preferred embodiment the target sequence is a codon, optionally more codons. The sunrise primer's sequence is intended to be duplicated by the nascent complementary strand and, in this way, the stem is destabilised, the two fluorophores are held
10 apart, usually between 15 and 25 nucleotides, and the fluorophore is free to emit its excitation energy for monitoring. The Scorpion primer resembles the sunrise primer, but derivate in having a moiety that blocks duplication on the signalling portion of the scorpion primer. The blocking moiety is typically hexethylene glycol. In addition to the difference in structure, the function of the scorpion primers differs slightly in that
15 the 5' region of the oligonucleotide is designed to hybridise to a complementary region within the amplicons. In a certain embodiment the complementary region is a codon on the identifier oligonucleotide. The hybridisation forces the labels apart disrupting the hairpin and permitting emission in the same way as the hairpin probes.

20 After the selection has been performed the codon profile is indicative of the chemical entities that have been used in the synthesis of encoded molecules having a certain property, such as an affinity towards a target. In the event the selection has been sufficient effective it may be possible directly to deduce a part or the entire structure of encoded molecules with the desired property. Alternatively, it may be possible to
25 deduce a structural unit appearing more frequently among the encoded molecules after the selection, which gives important information to the structure-activity-relationship (SAR). If the selection process has not narrowed the size of the library to a manageable number, the formation of a second-generation library is useful. In the formation of the second-generation library chemical entities, which have not
30 been involved in the synthesis of encoded molecules that have been successful in the selection may be omitted, thus limiting the size of the new library and at the same time increasing the concentration of complexes with the requested property, e.g. the ability to bind to a target. The second-generation library may then be subjected to more stringent selection conditions to allow only the encoded molecules
35 with a higher affinity to bind to the target. The second-generation library may also be

generated using the chemical entities coded for in addition to certain chemical entities suspected of increasing the performance of the final encoded molecule. The indication of certain successful chemical entities may be obtained from the SAR. The use in a second-generation library of chemical entities, which have proved to be
5 interesting for further investigation in a preceding library, may thus entail a shuffling with new chemical entities that may focus the second-generation library in a certain desired direction.

10 An Example of implicit identification of codons includes that the nucleic acid fragment is associated with a chemical entity precursor capable of being transferred to a recipient reactive group. The recipient reactive group may be a part of a chemical scaffold and the chemical entity precursor may add a structural unit to said scaffold. It is preferred that the nucleic acid fragment codes for the chemical entity. In some
15 comprises an anticodon, which identifies the chemical entity. When a plurality of chemical entities are present the anticodon is preferably unique, i.e. a unique correspondence between the chemical entities and the associated anticodons exists.

The identifier nucleic acid sequence comprises codons, which may be able to pair
20 with one or more anticodons of the pool of nucleic acid fragments. The pairing between one or more codons of an identifier nucleic acid sequence and one or more anticodons is preferably specific, i.e. the one or more codons of the identifier nucleic acid sequence are only recognized by particular anticodons. The nucleic acid fragment containing more than one anticodon can encode for scaffold molecules where
25 each anticodon encodes for specific chemical entities of that scaffold molecule. The specific pairing makes it possible implicitly to decode the codon of an identifier nucleic acid sequence. In the method according to the invention, non-specific pairing between codons and anticodons can be cleaved with an enzyme or chemically
30 treated to break the double stranded nucleotides. The non-pairing region can be cleaved using enzymes that cleaves specifically nucleotide sequences with mismatches. Notably, the enzyme is selected from T4 endonuclease VII, T4 endonuclease I, CEL I, nuclease S1, or variants thereof. The cleavage is preferable used when more than one codon and anticodon is involved in pairing between the identifier nucleic acid sequence and the nucleic acid fragment.

The pool of nucleic acid fragments associated with a chemical entity may comprise anticodons complemented by codons of one or more identifier nucleic acid sequence as well as anticodons which are not complemented by codons on any identifier nucleic acid sequence. In other words, the amount of genetic information contained in the anticodons of the pool is larger than the amount of genetic information complemented by the codons.

The contacting of the one or more identifier nucleic acid sequences with the pool of nucleic acid fragments are usually conducted at conditions, which allow for hybridisation, i.e. conditions at which cognate nucleic acid sequences can anneal to each other. To facilitate the recovery of nucleic acid fragments, which have annealed to the identifier nucleic acid sequences, the identifier nucleic acid sequences are usually immobilized on a solid support. Examples of suitable solid supports include beads and column material, e.g. beads and column material associated with a second part of the affinity pair to bind identifier nucleic acid sequences attached to the first part of the molecular affinity pair. In certain aspects of the invention the solid support is associated with streptavidin and the identifier nucleic acid sequences are attached to biotin.

When the identifier nucleic acid sequences are immobilized on a solid support the pool of nucleic acid fragments is typically present in a mobile phase, i.e. dissolved in a liquid. The identifier nucleic acids will hybridise to these nucleic acid fragments in the pool which are sufficient complementary to a particular part of an identifier nucleic acid sequence for a binding to occur. Fragments not finding any complementing sequence will remain in the solution. In the event, the identifier nucleic acid sequences are segregated into codons and the fragments comprises anticodons, the anticodons which are able to anneal to a codons will be caught while fragments not having a cognate codon will be maintained in the mobile phase. When codons and anticodons are present in the method of the present invention, specific hybridisation implies that the tendency of an anticodon to cross-hybridise to another codon will be impede or avoided. To avoid cross-hybridisation, codons may be designed such that each codon is distinguished from all other codons by one, two or more mismatching nucleotides.

The mobile phase is subsequently separated from the solid phase e.g. by washing, and the enriched pool of fragments is recovered. The recovery of the nucleic acid fragments are usually done by subjecting the hybrid to denaturing conditions, i.e. conditions which separate the two strands. If the parent nucleic acid sequences are
5 immobilized on beads, the separation of the fragments can be effected using denaturing conditions and centrifugation/spinning.

The enriched pool of nucleic acid fragments associated with a chemical entity may be used directly to prepare a next generation library of complexes, in which each
10 member of the library comprises an encoded molecule and the nucleic acid sequence which codes for this molecule. In one embodiment of the invention, building blocks comprising a particular transferable chemical entity associated with an anticodon corresponding to the anticodons of the detected fragments are used in the generation of the next generation library. In another embodiment, additional building
15 blocks are added having modified transferable chemical entities in order to improve on a certain property of the encoded molecule.

The complexes may be prepared by various known methods starting from the nucleic acid fragment comprising the anticodon and the chemical entity, as
20 disclosed above. According to a particular method, the next generation library is formed by a) mixing under hybridisation conditions, nascent bifunctional complexes comprising a chemical entity or a reaction product of chemical entities, and an identifier nucleic acid sequence comprising codon(s) identifying said chemical entities, with the recovered nucleic acid fragments, said fragments comprising an
25 oligonucleotide sufficient complementary to at least a part of the identifier nucleic acid sequence to allow for hybridisation, a transferable chemical entity and an anticodon identifying the chemical entity, to form hybridisation products; and b) transferring the chemical entities of the nucleic acid fragments to the nascent bifunctional complexes through a reaction involving a reactive group of the nascent
30 bifunctional complex, in conjunction with a transfer of the genetic information of the anticodon.

Preferably, the above method for preparing the next generation library comprises the further step of c) separating the components of the hybridisation product and
35 recovering the complexes. If further chemical entities are intended to participate in

the formation of the encoded molecule of the nascent complex, steps a) through c) are repeated as appropriate using the recovered complexes in step c) as the nascent bifunctional complexes in step a) of the next round.

- 5 The genetic information of the anticodon may be transferred to the nascent complex by a variety of methods. According to a first embodiment the genetic information of the anticodon is transferred by enzymatically extending the oligonucleotide identifier region to obtain a codon attached to the bifunctional complex having received the chemical entity. A second embodiment implies that genetic information of the anti-
- 10 codon is transferred to the nascent complexes by hybridisation to a cognate codon of the nascent complex.

According to the first embodiment, the enriched pool of fragments comprises an affinity oligonucleotide sufficient complementary to an identifier region of the nascent

15 complex, said oligonucleotide being distinct from the anticodon. Accordingly, the oligonucleotide identifier region of the nascent complex anneals to the affinity oligonucleotide of the building block to form the hybridisation product, while the anticodon remains single stranded. Subsequently, the chemical entity is transferred to the recipient reactive group of the complex to form the encoded molecule prior to,

20 simultaneously with, or subsequent to the enzymatically extension of the hybridisation product using the anticodon as identifier. Specific examples of suitable enzymes are polymerases and ligases, which requires dNTPs and oligonucleotides, respectively as substrates. The method for forming the complexes according to this first embodiment is the subject PCT/DK03/00739, the content thereof being incorporated

25 herein by reference.

According to the second embodiment, the anticodon form part of the affinity oligonucleotide, i.e. the anticodon is a part of or the entire affinity oligonucleotide. Initially, a plurality of identifiers comprising different codons and/or different order of codons is

30 provided. The identifiers are associated with a recipient reactive group, i.e. the reactive group may be covalently attached to the identifier or attached by hybridisation. Notably, a codon of the identifier may be used for the attachment of a building block harbouring the reactive group. The identifiers are subsequently contacted with the enriched pool of building blocks, i.e. nucleic acid fragments associated with a trans-

35 ferable chemical entity. The mixture of identifiers and building blocks are maintained

at hybridisation conditions to anneal the anticodon of the building blocks to the cognate codon of the identifier. After or simultaneously with the annealing step, the chemical entity is transferred to the recipient reactive group of the identifier. The method for forming the complexes according to the second embodiment is the subject of various patent applications, including WO 02/103008, WO 02/074929, Danish patent application No. PA 2002 01347, and US provisional patent application No. 60/409,968. The content of these patent applications are incorporated herein by reference in their entirety.

- 10 The new generation of library complexes may be used in a partition step, in which the library of complexes is subjected to a condition partitioning complexes displaying a predetermined property from the remainder of the next generation library, as explained above. Thus, using the present method, it is possible to repeat the partitioning procedure a desired number of times using still more stringent conditions, until a
- 15 single or a few encoded molecules are identified which display the desired property to a high extent. When the partitioning is based on an affinity assay, the library of encoded molecules are increasingly narrowed in size from one generation to the next and at the same time the high affinity binders are increased in concentration.
- 20 The outcome of a codon analysis will be dependent of the enrichment factor in the selection process. An efficient and specific selection will generate a large difference between the specific binders compared to the background. Still, there will be a large amount of molecules in the background that will reduce the possibility to obtain measurable differences between the binders and the background in the codon
- 25 analysis procedure. If the enrichment factor (or too large library) is not good enough to distinguish a specific binder among the background binders, the signal in the codon analysis will probably not be detectable. However, there will be a continuing of binders that use a certain chemical entity in a certain position. These "non optimal" binders (a certain important chemical entity in one position and less important
- 30 in the other position) will be many due to the diversity obtained when only one (or a few) positions are important in the selection process. Therefore, the sum of all molecules with a preferable chemical entity in a certain position will be larger than the sum of all molecules with a non-binding chemical entity, which will make the codon analysis easier.

This invention may involve an extensive analysis of all the chemical entities in a library and how they are involved in the binding to targets. This information can be used both to design new libraries and in the final process where the lead structures are produced and pre-clinical candidates are picked. The extensive data obtained in the codon analysis can for instance be used for selecting candidates with the appropriate specificity. This can be done if selection has been performed on a family of proteins where one of the members is the target.

The invention enables pharmacophore identification and transformation into small molecule drugs. In cases where peptide-like libraries is used, the peptide/petidomimetic lead to small molecule conversion process is supported by medicinal chemistry and cheminformatics and guided by matching the pharmacophore derived from massive structure activity relationship (SAR) data information from the codon analysis. A "pharmacophore" is a description of the structural criteria a molecule must fulfil in order that it is active against a specified biological receptor. These criteria are usually the 3D spatial relationships of a set of chemical features, and sometimes include the steric boundaries, within which the molecule must fit. There is a set of software methods, which automatically infers such pharmacophores, given a SAR, in the absence of direct macromolecular structural data.

The extensive SAR information obtained using the codon analyses described in this invention can be combined with molecular modeling technologies to refine for example pharmacophore models and the plausible interactions between the potential binders and a target.

The codon analysis is also a valuable experimental tool for SAR on weak binders. The codon analysis measures the abundance of chemical entities after a selection in all binding molecules. Thus, even week binders, which there might be many of, is detected even though the detected codon is selected in many different combinations. The selection procedure can also be tuned to enrich predominately for weak binders, which will simplify the codon analysis data.

This invention is also suitable for replacing the laborious task of extracting SAR information by hand with an automated process using suitable algorithm and software programs. The codon analysis (e.g. array or QPCR measurements) can be directly

feed into a data handling software program that use both the codon abundances and structural data to generate SAR information and potential pharmacophore models.

- 5 The SAR information and potential pharmacophore models obtained from the codon analysis can be used to design focused libraries in an array format allowing massive and parallel testing. Thus, the selection procedure and codon analysis can be seen as a diversity reduction step to allow a complete test of potential binders in an array format.

10

Various methods for identifying the codons of the identifiers of step iii) are disclosed herein. When a pool of partitioned identifier nucleic acid sequences is subjected to the identification step it is normally not practically to decode a sufficient number of sequences comprising the entire "genome" of an encoded molecule to ensure that
15 all interesting encoded molecules have been revealed. Therefore, a modified sequencing technique preferably identifies the codons in each position occurring with the highest frequency. The next generation library is then build using in each position the chemical entities occurring with the highest frequency.

- 20 In a certain embodiment of the invention, the codon identification step uses the entire population of identifier nucleic acid sequences in the analysis and informs the experimenter of the relative abundance of each codon in a certain position. The codon information may be obtained using microarray, QPCR, or any equivalent method for revealing the identity of codons. In contrary, sequencing a subset of
25 identifier nucleic acid sequences only provides the experimenter with a limited insight as to the population of codons and the corresponding encoded molecules.

Detailed Description of the Invention

Complex

- 30 The complex comprises an encoded molecule and an identifier oligonucleotide. The identifier comprises codons that identify the encoded molecule. Preferably, the identifier oligonucleotide identifies the encoded molecule uniquely, i.e. in a library of complexes a particular identifier is capable of distinguishing the molecule it is attached to from the rest of the molecules.

35

The encoded molecule and the identifier may be attached directly to each other or through a bridging moiety. In one aspect of the invention, the bridging moiety is a selectively cleavable linkage.

- 5 The identifier oligonucleotide may comprise two or more codons. In a preferred aspect the identifier oligonucleotide comprises three or more codons. The sequence of each codon can be decoded utilizing the present method to identify reactants used in the formation of the encoded molecule. When the identifier comprises more than one codon, each member of a pool of chemical entities can be identified and the
10 order of codons is informative of the synthesis step each member has been incorporated in.

In a certain embodiment, the same codon is used to code for several different chemical entities. In a subsequent identification step, the structure of the encoded
15 molecule can be deduced taking advantage of the knowledge of different attachment chemistries, steric hindrance, deprotection of orthogonal protection groups, etc. In another embodiment, the same codon is used for a group of chemical entities having a common property, such as a lipophilic nature, a certain attachment chemistry etc. In a preferred embodiment, however, the codon is unique i.e. a similar combination of nucleotides does not appear on the identifier oligonucleotide coding for another chemical entity. In a practical approach, for a specific chemical entity, only a single combination of nucleotides is used. In some aspects of the invention, it may be advantageous to use several codons for the same chemical entity, much in the same way as Nature uses up to six different codons for a single amino acid. The two
20 or more codons identifying the same chemical entity may carry further information related to different reaction conditions.

The sequence of the nucleotides in each codon may have any suitable length. The codon may be a single nucleotide or a plurality of nucleotides. In some aspects of
30 the invention, it is preferred that each codon independently comprises four or more nucleotides, more preferred 4 to 30 nucleotides. In some aspects of the invention the lengths of the codons vary.

A certain codon may be distinguished from any other codon in the library by only a
35 single nucleotide. However, to facilitate a subsequent decoding process and to in-

crease the ability of the primer to discriminate between codons it is in general desired to have two or more mismatches between a particular codon and any other codon appearing on identifier oligonucleotide. As an example, if a codon length of 5 nucleotides is selected, more than 100 nucleotide combinations exist in which two or more mismatches appear. For a certain number of nucleotides in the codon, it is generally desired to optimize the number of mismatches between a particular codon relative to any other codon appearing in the library.

The identifier oligonucleotide will in general have at least two codons arranged in sequence, i.e. next to each other. Two neighbouring codons may be separated by a framing sequence. Depending on the encoded molecule formed, the identifier may comprise further codons, such as 3, 4, 5, or more codons. Each of the further codons may be separated by a suitable framing sequence. Preferably, all or at least a majority of the codons of the identifier are separated from a neighbouring codon by a framing sequence. The framing sequence may have any suitable number of nucleotides, e.g. 1 to 20. Alternatively, codons on the identifier may be designed with overlapping sequences.

The framing sequence, if present, may serve various purposes. In one setup of the invention, the framing sequence identifies the position of the codon. Usually, the framing sequence either upstream or downstream of a codon comprises information which positions the chemical entity and the reaction conditions in the synthesis history of the encoded molecule. The framing sequence may also or in addition provide for a region of high affinity. The high affinity region may ensure that a hybridisation event with an anti-codon will occur in frame. Moreover, the framing sequence may adjust the annealing temperature to a desired level.

A framing sequence with high affinity can be provided by incorporation of one or more nucleobases forming three hydrogen bonds to a cognate nucleobase. Examples of nucleobases having this property are guanine and cytosine. Alternatively, or in addition, the framing sequence may be subjected to backbone modification. Several back bone modifications provides for higher affinity, such as 2'-O-methyl substitution of the ribose moiety, peptide nucleic acids (PNA), and 2'-4' O-methylene cyclisation of the ribose moiety, also referred to as LNA (Locked Nucleic Acid).

The sequence comprising a codon and an adjacent framing sequence has in a certain aspect of the invention a total length of 11 nucleotides or more, preferably 15 nucleotides or more. A primer may be designed to complementary to the codon sequence as well as the framing sequence. The presence of an extension reaction
5 under conditions allowing for such reaction to occur is indicative of the presence of the chemical entity encoded in the codon as well as the position said chemical entity has in the entire synthesis history of the encoded molecule.

The identifier may comprise flanking regions around the coding section. The flanking
10 regions can also serve as priming sites for amplification reactions, such as PCR or as binding region for oligonucleotide probe. The identifier may in certain embodiments comprise an affinity region having the property of being able to hybridise to a building block.

15 It is to be understood that when the term identifier oligonucleotide is used in the present description and claims, the identifier oligonucleotide may be in the sense or the anti-sense format; i.e. the identifier can be a sequence of codons which actually codes for the encoded molecule or can be a sequence complementary thereto. Moreover, the identifier may be single-stranded or double-stranded, as appropriate.

20 The encoded molecule part of the complex is generally of a structure expected of having an effect according to the property sought for, e.g. the encoded molecule has a binding affinity towards a target. When the target is of pharmaceutical importance, the encoded molecule is generally a possible drug candidate. The complex may be
25 formed by tagging a library of different possible drug candidates with a tag, e.g. a nucleic acid tag identifying each possible drug candidate. In another embodiment of the invention, the molecule formed by a variety of reactants which have reacted with each other and/or a scaffold molecule. Optionally, this reaction product may be post-modified to obtain the final molecule displayed on the complex. The post-
30 modification may involve the cleavage of one or more chemical bonds attaching the encoded molecule to the identifier in order more efficiently to display the encoded molecule.

The formation of an encoded molecule generally starts by a scaffold, i.e. a chemical
35 unit having one or more reactive groups capable of forming a connection to another

reactive group positioned on a chemical entity, thereby generating an addition to the original scaffold. A second chemical entity may react with a reactive group also appearing on the original scaffold or a reactive group incorporated by the first chemical entity. Further chemical entities may be involved in the formation of the

5 final reaction product. The formation of a connection between the chemical entity and the nascent encoded molecule may be mediated by a bridging molecule. As an example, if the nascent encoded molecule and the chemical entity both comprise an amine group a connection between these can be mediated by a dicarboxylic acid. A synthetic molecule is in general produced in vitro and may be a naturally occurring

10 or an artificial substance. Usually, a synthetic molecule is not produced using the naturally translation system in an in vitro process.

The chemical entities that are precursors for structural additions or eliminations of the encoded molecule may be attached to a building block prior to the participation

15 in the formation of the reaction product leading the final encoded molecule. Besides the chemical entity, the building block generally comprises an anti-codon. In some embodiments the building blocks also comprise an affinity region providing for affinity towards the nascent complex.

20 Thus, the chemical entities are suitably mediated to the nascent encoded molecule by a building block, which further comprises an anticodon. The anti-codon serves the function of transferring the genetic information of the building block in conjunction with the transfer of a chemical entity. The transfer of genetic information and chemical entity may occur in any order. The chemical entities are preferably

25 reacted without enzymatic interaction in some aspects of the invention. Notably, the reaction of the chemical entities is preferably not mediated by ribosomes or enzymes having similar activity. In other aspects of the invention, enzymes are used to mediate the reaction between a chemical entity and a nascent encoded molecule.

30 According to certain aspects of the invention the genetic information of the anti-codon is transferred by specific hybridisation to a codon on a nucleic acid identifier. Another method for transferring the genetic information of the anti-codon to the nascent complex is to anneal an oligonucleotide complementary to the anti-codon and attach this oligonucleotide to the complex, e.g. by ligation. A still further method

involves transferring the genetic information of the anti-codon to the nascent complex by an extension reaction using a polymerase and a mixture of dNTPs.

5 The chemical entity of the building block may in most cases be regarded as a precursor for the structural entity eventually incorporated into the encoded molecule. In other cases the chemical entity provides for the eliminations of chemical units of the nascent encoded molecule. Therefore, when it in the present application with claims is stated that a chemical entity is transferred to a nascent encoded molecule it is to be understood that not necessarily all the atoms of the original chemical entity is to
10 be found in the eventually formed encoded molecule. Also, as a consequence of the reactions involved in the connection, the structure of the chemical entity can be changed when it appears on the nascent encoded molecule. Especially, the cleavage resulting in the release of the entity may generate a reactive group which in a subsequent step can participate in the formation of a connection between a nascent
15 complex and a chemical entity.

The chemical entity of the building block comprises at least one reactive group capable of participating in a reaction which results in a connection between the chemical entity of the building block and another chemical entity or a scaffold associated
20 with the nascent complex. The number of reactive groups which appear on the chemical entity is suitably one to ten. A building block featuring only one reactive group is used i.a. in the end positions of polymers or scaffolds, whereas building blocks having two reactive groups are suitable for the formation of the body part of a polymer or scaffolds capable of being reacted further. One, two or more reactive
25 groups intended for the formation of connections, are typically present on scaffolds. Non-limiting examples of scaffolds are opiates, steroids, benzodiazepines, hydantoines, and peptidylphosphonates.

The reactive group of the chemical entity may be capable of forming a direct connection to a reactive group of the nascent complex or the reactive group of the building block may be capable of forming a connection to a reactive group of the nascent complex through a bridging fill-in group. It is to be understood that not all the atoms of a reactive group are necessarily maintained in the connection formed. Rather, the reactive groups are to be regarded as precursors for the structure of the connection.

The subsequent cleavage step to release the chemical entity from the building block can be performed in any appropriate way. In an aspect of the invention the cleavage involves usage of a chemical reagent or an enzyme. The cleavage results in a transfer of the chemical entity to the nascent encoded molecule or in a transfer of the nascent encoded molecule to the chemical entity of the building block. In some cases it may be advantageous to introduce new chemical groups as a consequence of linker cleavage. The new chemical groups may be used for further reaction in a subsequent cycle, either directly or after having been activated. In other cases it is desirable that no trace of the linker remains after the cleavage.

10 In another aspect, the connection and the cleavage is conducted as a simultaneous reaction, i.e. either the chemical entity of the building block or the nascent encoded molecule is a leaving group of the reaction. In some aspects of the invention, it is appropriate to design the system such that the connection and the cleavage occur simultaneously because this will reduce the number of steps and the complexity. The simultaneous connection and cleavage can also be designed such that either no trace of the linker remains or such that a new chemical group for further reaction is introduced, as described above.

20 The attachment of the chemical entity to the building block, optionally via a suitable spacer can be at any entity available for attachment, e.g. the chemical entity can be attached to a nucleobase or the backbone. In general, it is preferred to attach the chemical entity at the phosphor of the internucleoside linkage or at the nucleobase. When the nucleobase is used for attachment of the chemical entity, the attachment point is usually at the 7 position of the purines or 7-deaza-purins or at the 5 position of pyrimidines. The nucleotide may be distanced from the reactive group of the chemical entity by a spacer moiety. The spacer may be designed such that the conformational space sampled by the reactive group is optimized for a reaction with the reactive group of the nascent encoded molecule.

30 The encoded molecules may have any chemical structure. In a preferred aspect, the encoded molecule can be any compound that may be synthesized in a component-by-component fashion. In some aspects the synthetic molecule is a linear or branched polymer. In another aspect the synthetic molecule is a scaffolded molecule. The term "encoded molecule" also comprises naturally occurring

35

molecules like α -polypeptides etc, however produced *in vitro* usually in the absence of enzymes, like ribosomes. In certain aspects, the synthetic molecule of the library is a non- α -polypeptide.

- 5 The encoded molecule may have any molecular weight. However, in order to be orally available, it is in this case preferred that the synthetic molecule has a molecular weight less than 2000 Daltons, preferably less than 1000 Dalton, and more preferred less than 500 Daltons.
- 10 The size of the library may vary considerably pending on the expected result of the inventive method. In some aspects, it may be sufficient that the library comprises two, three, or four different complexes. However, in most events, more than two different complexes are desired to obtain a higher diversity. In some aspects, the library comprises 1,000 or more different complexes, more preferred 1,000,000 or
- 15 more different complexes. The upper limit for the size of the library is only restricted by the size of the vessel in which the library is comprised. It may be calculated that a vial may comprise up to 10^{14} different complexes.

Methods for forming libraries of complexes

- 20 The encoded molecules associated with an identifier oligonucleotide having two or more codons that code for reactants that have reacted in the formation of the molecule part of the complex may be formed by a variety of processes. Generally, the preferred methods can be used for the formation of virtually any kind of encode molecule. Suitable examples of processes include prior art methods disclosed in
- 25 WO 93/20242, WO 93/06121, WO 00/23458, WO 02/074929, and WO 02/103008, the content of which being incorporated herein by reference as well as methods of the present applicant not yet public available, including the methods disclosed in PCT/DK03/00739 filed 30 October 2003, and DK PA 2003 00430 filed 20 March 2003. Any of these methods may be used, and the entire content of the patent ap-
- 30 plications are included herein by reference.

- Below five presently preferred embodiments are described. A first embodiment disclosed in more detail in WO 02/103008 is based on the use of a polymerase to incorporate unnatural nucleotides as building blocks. Initially, a plurality of identifier
- 35 oligonucleotides is provided. Subsequently primers are annealed to each of the

identifiers and a polymerase is extending the primer using nucleotide derivatives, which have appended chemical entities. Subsequent to or simultaneously with the incorporation of the nucleotide derivatives, the chemical entities are reacted to form a reaction product. The encoded molecule may be post-modified by cleaving some
5 of the linking moieties to better present the encoded molecule.

Several possible reaction approaches for the chemical entities are apparent. First, the nucleotide derivatives can be incorporated and the chemical entities subsequently polymerised. In the event the chemical entities each carry two reactive
10 groups, the chemical entities can be attached to adjacent chemical entities by a reaction of these reactive groups. Exemplary of the reactive groups are amine and carboxylic acid, which upon reaction form an amide bond. Adjacent chemical entities can also be linked together using a linking or bridging moiety. Exemplary of this
15 approach is the linking of two chemical entities each bearing an amine group by a bi-carboxylic acid. Yet another approach is the use of a reactive group between a chemical entity and the nucleotide building block, such as an ester or a hoister group. An adjacent building block having a reactive group such as an amine may cleave the interspaced reactive group to obtain a linkage to the chemical entity, e.g.
by an amide linking group.

20 A second embodiment for obtainment of complexes disclosed in WO 02/103008 pertains to the use of hybridisation of building blocks to an identifier and reaction of chemical entities attached to the building blocks in order to obtain a reaction product. This approach comprises that identifiers are contacted with a plurality of
25 building blocks, wherein each building block comprises an anti-codon and a chemical entity. The anti-codons are designed such that they recognise a sequence, i.e. a codon, on the identifier. Subsequent to the annealing of the anti-codon and the codon to each other a reaction of the chemical entity is effected.

30 The identifier may be associated with a scaffold. Building blocks bringing chemical entities in may be added sequentially or simultaneously and a reaction of the reactive group of the chemical entity may be effected at any time after the annealing of the building blocks to the identifier.

A third embodiment for the generation of a complex includes chemical or enzymatic ligation of building blocks when these are lined up on a identifier. Initially, identifiers are provided, each having one or more codons. The identifiers are contacted with building blocks comprising anti-codons linked to chemical entities. The two or more
5 anti-codons annealed on an identifier are subsequently ligated to each other and a reaction of the chemical entities is effected to obtain a reaction product. The method is disclosed in more detail in DK PA 2003 00430 filed 20 March 2003.

A fourth embodiment makes use of the extension by a polymerase of an affinity
10 sequence of the nascent complex to transfer the anti-codon of a building block to the nascent complex. The method implies that a nascent complex comprising a scaffold and an affinity region is annealed to a building block comprising a region complementary to the affinity section. Subsequently, the anti-codon region of the building block is transferred to the nascent complex by a polymerase. The transfer
15 of the chemical entity may be transferred prior to, simultaneously with or subsequent to the transfer of the anti-codon. This method is disclosed in detail in PCT/DK03/00739.

A fifth embodiment also disclosed in PCT/DK03/00739 comprises reaction of a
20 reactant with a reaction site on nascent bifunctional molecule and addition of a nucleic acid tag to the nascent bifunctional molecule using an enzyme, such as a ligase. When a library is formed, usually an array of compartments is used for reaction of reactants and enzymatic addition of tags with the nascent bifunctional molecule.

25

Thus, the codons are either pre-made into one or more identifiers before the encoded molecules are generated or the codons are transferred simultaneously with the formation of the encoded molecules.

30 After or simultaneously with the formation of the reaction product some of the linkers to the identifier may be cleaved, however, usually at least one linker is maintained to provide for the complex.

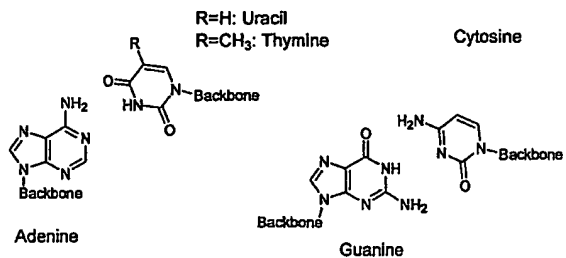
Nucleotides

The nucleotides used in the present invention may be linked together in a sequence of nucleotides, i.e. an oligonucleotide. Each nucleotide monomer is normally composed of two parts, namely a nucleobase moiety, and a backbone. The backbone
5 may in some cases be subdivided into a sugar moiety and an internucleoside linker.

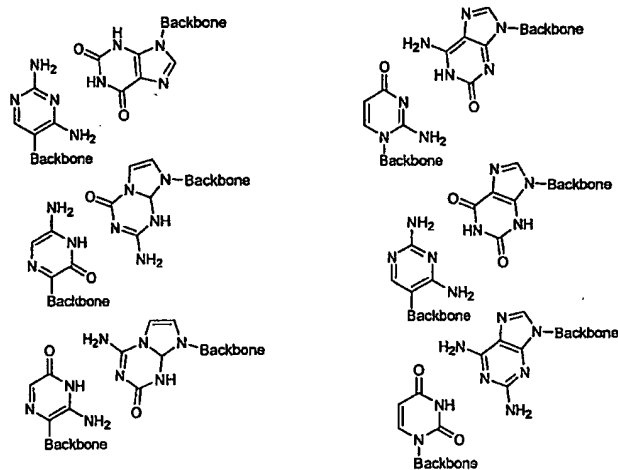
The nucleobase moiety may be selected among naturally occurring nucleobases as well as non-naturally occurring nucleobases. Thus, "nucleobase" includes not only
10 the known purine and pyrimidine hetero-cycles, but also heterocyclic analogues and tautomers thereof. Illustrative examples of nucleobases are adenine, guanine, thymine, cytosine, uracil, purine, xanthine, diaminopurine, 8-oxo-N⁶-methyladenine, 7-deazaxanthine, 7-deazaguanine, N⁴,N⁴-ethanocytosin, N⁶,N⁶-ethano-2,6-diaminopurine, 5-methylcytosine, 5-(C³-C⁶)-alkynylcytosine, 5-fluorouracil, 5-bromouracil,
15 pseudoisocytosine, 2-hydroxy-5-methyl-4-triazolopyridine, isocytosine, isoguanine, inosine and the "non-naturally occurring" nucleobases described in Benner et al., U.S. Pat No. 5,432,272. The term "nucleobase" is intended to cover these examples as well as analogues and tautomers thereof. Especially interesting nucleobases are adenine, guanine, thymine, cytosine, 5-methylcytosine, and uracil, which are con-
20 sidered as the naturally occurring nucleobases in relation to therapeutic and diagnostic application in humans.

Examples of suitable specific pairs of nucleobases are shown below:

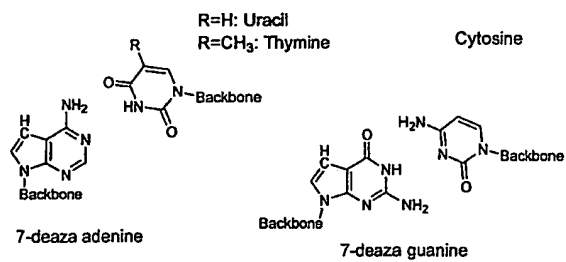
Natural Base Pairs



Synthetic Base Pairs

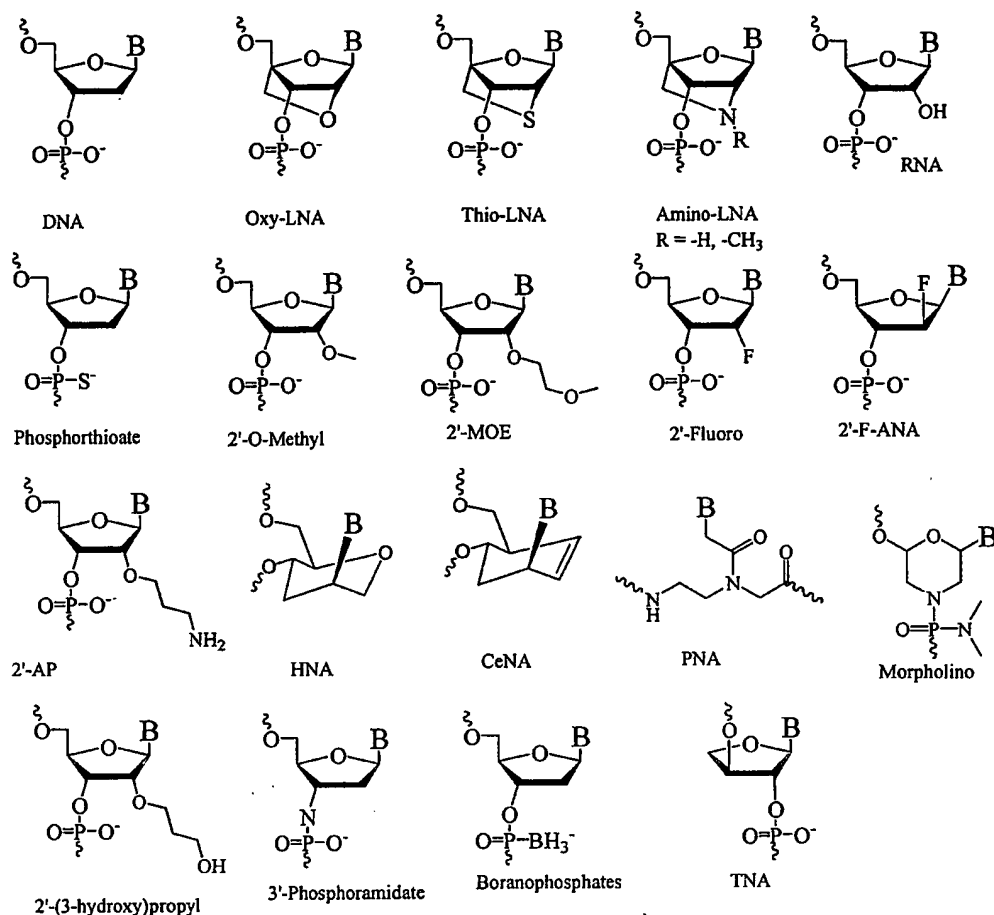


Synthetic purine bases pairing with natural pyrimidines



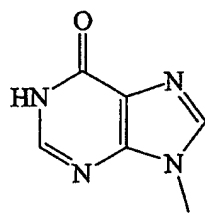
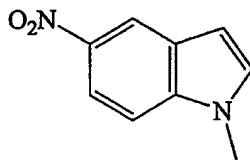
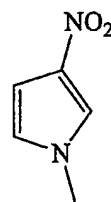
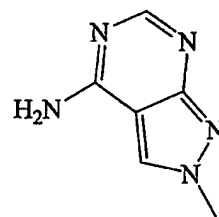
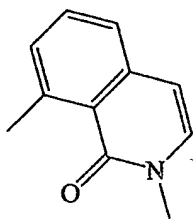
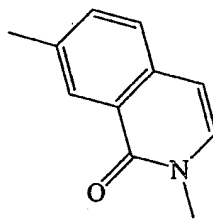
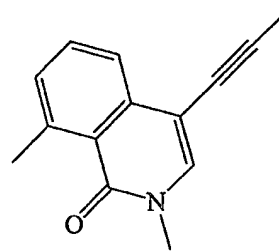
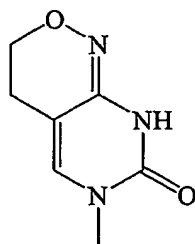
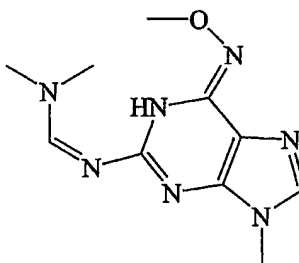
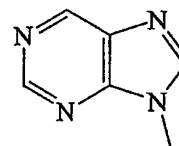
Suitable examples of backbone units are shown below (B denotes a nucleobase):

34



- The sugar moiety of the backbone is suitably a pentose but may be the appropriate part of a PNA or a six-member ring. Suitable examples of possible pentoses include ribose, 2'-deoxyribose, 2'-O-methyl-ribose, 2'-fluor-ribose, and 2'-4'-O-methylene-ribose (LNA). Suitably the nucleobase is attached to the 1' position of the pentose entity.
- 5
- 10 An internucleoside linker connects the 3' end of preceding monomer to a 5' end of a succeeding monomer when the sugar moiety of the backbone is a pentose, like ribose or 2-deoxyribose. The internucleoside linkage may be the natural occurring phosphodiester linkage or a derivative thereof. Examples of such derivatives include phosphorothioate, methylphosphonate, phosphoramidate, phosphotriester, and
- 15 phosphodithioate. Furthermore, the internucleoside linker can be any of a number of non-phosphorous-containing linkers known in the art.

Preferred nucleic acid monomers include naturally occurring nucleosides forming part of the DNA as well as the RNA family connected through phosphodiester linkages. The members of the DNA family include deoxyadenosine, deoxyguanosine, 5 deoxythymidine, and deoxycytidine. The members of the RNA family include adenosine, guanosine, uridine, cytidine, and inosine. Inosine is a non-specific pairing nucleoside and may be used as universal base because inosine can pair nearly isoenergetically with A, T, and C. Other compounds having the same ability of non-specifically base-pairing with natural nucleobases have been formed. Suitable compounds which may be utilized in the present invention includes among others the 10 compounds depicted below

Examples of Universal Bases:**Inosine****5-Nitroindole****3-Nitropyrrole****N⁸-8aza-7deazaadenine****MICS****5MICS****PIM****dP****dK****Nebularine**

Building block

The chemical entities or reactants that are precursors for structural additions or eliminations of the encoded molecule may be attached to a building block prior to the participation in the formation of the reaction product leading to the final encoded molecule. Besides the chemical entity, the building block generally comprises an anti-codon.

The chemical entity of the building block comprises at least one reactive group capable of participating in a reaction, which results in a connection between the chemical entity of the building block and another chemical entity or a scaffold associated with the nascent complex. The connection is facilitated by one or more reactive groups of the chemical entity. The number of reactive groups, which appear on the chemical entity, is suitably one to ten. A building block featuring only one reactive group is used i.a. in the end positions of polymers or scaffolds, whereas building blocks having two reactive groups are suitable for the formation of the body part of a polymer or scaffolds capable of being reacted further. One, two or more reactive groups intended for the formation of connections are typically present on scaffolds.

The reactive group of the building block may be capable of forming a direct connection to a reactive group of the nascent complex or the reactive group of the building block may be capable of forming a connection to a reactive group of the nascent complex through a bridging fill-in group. It is to be understood that not all the atoms of a reactive group are necessarily maintained in the connection formed. Rather, the reactive groups are to be regarded as precursors for the structure of the connection.

The subsequent cleavage step to release the chemical entity from the building block can be performed in any appropriate way. In an aspect of the invention the cleavage involves usage of a reagent or an enzyme. The cleavage results in a transfer of the chemical entity to the nascent encoded molecule or in a transfer of the nascent encoded molecule to the chemical entity of the building block. In some cases it may be advantageous to introduce new chemical groups as a consequence of linker cleavage. The new chemical groups may be used for further reaction in a subsequent cycle, either directly or after having been activated. In other cases it is desirable that no trace of the linker remains after the cleavage.

In another aspect, the connection and the cleavage are conducted as a simultaneous reaction, i.e. either the chemical entity of the building block or the nascent encoded molecule is a leaving group of the reaction. In general, it is preferred to design the system such that the connection and the cleavage occur simultaneously because this will reduce the number of steps and the complexity. The simultaneous connection and cleavage can also be designed such that either no trace of the linker remains or such that a new chemical group for further reaction is introduced, as described above.

10

The attachment of the chemical entity to the building block, optionally via a suitable spacer can be at any entity available for attachment, e.g. the chemical entity can be attached to a nucleobase or the backbone. In general, it is preferred to attach the chemical entity at the phosphor of the internucleoside linkage or at the nucleobase. When the nucleobase is used for attachment of the chemical entity, the attachment point is usually at the 7 position of the purines or 7-deaza-purins or at the 5 position of pyrimidines. The nucleotide may be distanced from the reactive group of the chemical entity by a spacer moiety. The spacer may be designed such that the conformational space sampled by the reactive group is optimized for a reaction with the reactive group of the nascent encoded molecule or reactive site.

20

The anticodon complements the codon of the identifier oligonucleotide sequence and generally comprises the same number of nucleotides as the codon. The anticodon may be adjoined with a fixed sequence, such as a sequence complementing a framing sequence.

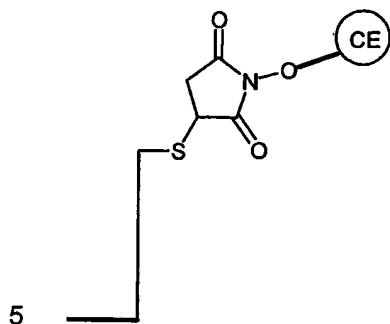
25

Various specific building blocks are envisaged. Building blocks of particular interest are shown below.

30 *Building blocks transferring a chemical entity to a recipient nucleophilic group*

The building block indicated below is capable of transferring a chemical entity (CE) to a recipient nucleophilic group, typically an amine group. The bold lower horizontal line illustrates the building block comprising an anti-codon and the vertical line illustrates a spacer. The 5-membered substituted N-hydroxysuccinimid (NHS) ring

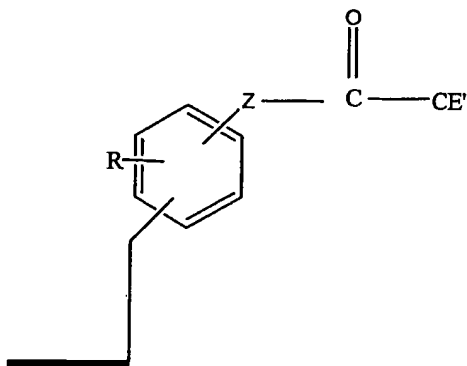
serves as an activator, i.e. a labile bond is formed between the oxygen atom connected to the NHS ring and the chemical entity. The labile bond may be cleaved by a nucleophilic group, e.g. positioned on a scaffold



The 5-membered substituted N-hydroxysuccinimid (NHS) ring serves as an activator, i.e. a labile bond is formed between the oxygen atom connected to the NHS ring and the chemical entity. The labile bond may be cleaved by a nucleophilic group, e.g. positioned on a scaffold, to transfer the chemical entity to the scaffold, thus converting the remainder of the fragment into a leaving group of the reaction.

When the chemical entity is connected to the activator through a carbonyl group and the recipient group is an amine, the bond formed on the scaffold will be an amide bond. The above building block is the subject of WO03078627A2, the content of which is incorporated herein in their entirety by reference.

Another building block, which may form an amide bond, is



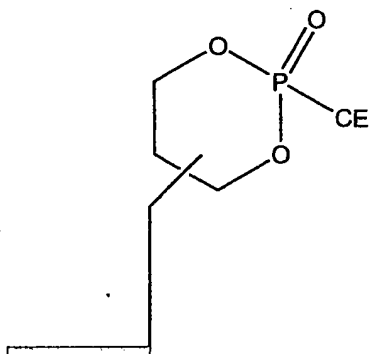
R may be absent or NO₂, CF₃, halogen, preferably Cl, Br, or I, and Z may be S or O. This type of building block is disclosed in WO03078626A2. The content of this patent application is incorporated herein in the entirety by reference.

A nucleophilic group can cleave the linkage between Z and the carbonyl group thereby transferring the chemical entity $-(C=O)-CE'$ to said nucleophilic group.

5 *Building blocks transferring a chemical entity to a recipient reactive group forming a C=C bond*

A building block as shown below is able to transfer the chemical entity to a recipient aldehyde group thereby forming a double bond between the carbon of the aldehyde and the chemical entity

10



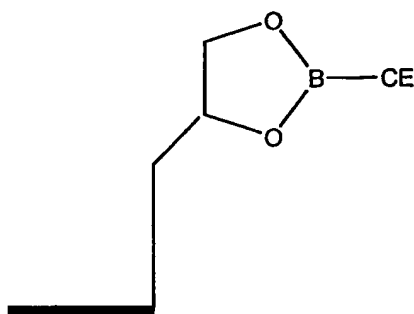
The above building block is disclosed in WO03078445A2, the content of which being incorporated herein in the entirety by reference.

15

Building blocks transferring a chemical entity to a recipient reactive group forming a C-C bond

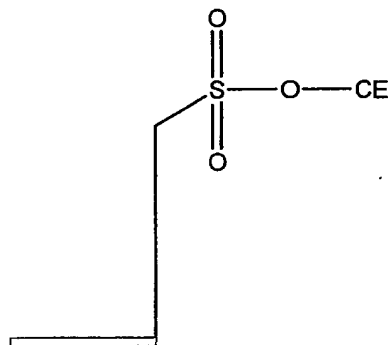
The below building block is able to transfer the chemical entity to a recipient group thereby forming a single bond between the receiving moiety, e.g. a scaffold, and the chemical entity.

20



The above building block is disclosed in WO03078445A2, the content of which being incorporated herein in the entirety by reference.

- 5 Another building block capable of transferring a chemical entity to a receiving reactive group forming a single bond is



- 10 The receiving group may be a nucleophile, such as a group comprising a hetero atom, thereby forming a single bond between the chemical entity and the hetero atom, or the receiving group may be an electronegative carbon atom, thereby forming a C-C bond between the chemical entity and the scaffold. The above building block is disclosed in WO03078446A2, the content of which is incorporated herein by
 15 reference.

- The chemical entity attached to any of the above building blocks may be a selected from a large arsenal of chemical structures. Examples of chemical entities are H or entities selected among the group consisting of a C₁-C₆ alkyl, C₂-C₆ alkenyl,
 20 C₂-C₆ alkynyl, C₄-C₈ alkadienyl, C₃-C₇ cycloalkyl, C₃-C₇ cycloheteroalkyl, aryl, and heteroaryl, said group being substituted with 0-3 R⁴, 0-3 R⁵ and 0-3 R⁹ or C₁-C₃ alkylene-NR⁴₂, C₁-C₃ alkylene-NR⁴C(O)R⁸, C₁-C₃ alkylene-NR⁴C(O)OR⁸, C₁-C₂ al-

kylene-O-NR⁴₂, C₁-C₂ alkylene-O-NR⁴C(O)R⁸, C₁-C₂ alkylene-O-NR⁴C(O)OR⁸ substituted with 0-3 R⁹.

where R⁴ is H or selected independently among the group consisting of C₁-C₆ alkyl, C₂-C₆ alkenyl, C₂-C₆ alkynyl, C₃-C₇ cycloalkyl, C₃-C₇ cycloheteroalkyl, aryl, heteroaryl, said group being substituted with 0-3 R⁹ and

R⁵ is selected independently from -N₃, -CNO, -C(NOH)NH₂, -NHOH, -NHNHR⁶, -C(O)R⁶, -SnR⁶₃, -B(OR⁶)₂, -P(O)(OR⁶)₂ or the group consisting of C₂-C₆ alkenyl, C₂-C₆ alkynyl, C₄-C₈ alkadienyl said group being substituted with 0-2 R⁷,

where R⁶ is selected independently from H, C₁-C₆ alkyl, C₃-C₇ cycloalkyl, aryl or C₁-C₆ alkylene-aryl substituted with 0-5 halogen atoms selected from -F, -Cl, -Br, and -I; and

R⁷ is independently selected from -NO₂, -COOR⁶, -COR⁶, -CN, -OSiR⁶₃, -OR⁶ and -NR⁶₂.

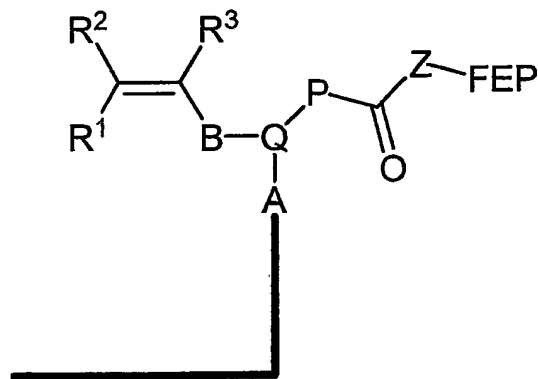
R⁸ is H, C₁-C₆ alkyl, C₂-C₆ alkenyl, C₂-C₆ alkynyl, C₃-C₇ cycloalkyl, aryl or C₁-C₆ alkylene-aryl substituted with 0-3 substituents independently selected from -F, -Cl, -NO₂, -R³, -OR³, -SiR³₃

R⁹ is =O, -F, -Cl, -Br, -I, -CN, -NO₂, -OR⁶, -NR⁶₂, -NR⁶-C(O)R⁸, -NR⁶-C(O)OR⁸, -SR⁶, -S(O)R⁶, -S(O)₂R⁶, -COOR⁶, -C(O)NR⁶₂ and -S(O)₂NR⁶₂.

20 Cross-link cleavage building blocks

It may be advantageous to split the transfer of a chemical entity to a recipient reactive group into two separate steps, namely a cross-linking step and a cleavage step because each step can be optimized. A suitable building block for this two-step process is illustrated below:

25



Initially, a reactive group appearing on the chemical entity precursor (abbreviated FEP) reacts with a recipient reactive group, e.g. a reactive group appearing on a scaffold, thereby forming a cross-link. Subsequently, a cleavage is performed, usually by adding an aqueous oxidising agent such as I₂, Br₂, Cl₂, H⁺, or a Lewis acid.

- 5 The cleavage results in a transfer of the group HZ-FEP- to the recipient moiety, such as a scaffold.

In the above formula

Z is O, S, NR⁴

- 10 Q is N, CR¹

P is a valence bond, O, S, NR⁴, or a group C₅₋₇arylene, C₁₋₆alkylene, C₁₋₆O-alkylene, C₁₋₆S-alkylene, NR¹-alkylene, C₁₋₆alkylene-O, C₁₋₆alkylene-S option said group being substituted with 0-3 R⁴, 0-3 R⁵ and 0-3 R⁹ or C₁₋₃ alkylene-NR⁴₂, C₁₋₃ alkylene-NR⁴C(O)R⁸, C₁₋₃ alkylene-NR⁴C(O)OR⁸, C₁₋₂ alkylene-O-NR⁴₂,
15 C₁₋₂ alkylene-O-NR⁴C(O)R⁸, C₁₋₂ alkylene-O-NR⁴C(O)OR⁸ substituted with 0-3 R⁹,

B is a group comprising D-E-F, in which

- 20 D is a valence bond or a group C₁₋₆alkylene, C₁₋₆alkenylene, C₁₋₆alkynylene, C₅₋₇arylene, or C₅₋₇heteroarylene, said group optionally being substituted with 1 to 4 group R¹¹,

E is, when present, a valence bond, O, S, NR⁴, or a group C₁₋₆alkylene, C₁₋₆alkenylene, C₁₋₆alkynylene, C₅₋₇arylene, or C₅₋₇heteroarylene, said group optionally being substituted with 1 to 4 group R¹¹,

F is, when present, a valence bond, O, S, or NR⁴,

- 25 A is a spacing group distancing the chemical structure from the complementing element, which may be a nucleic acid,

- 30 R¹, R², and R³ are independent of each other selected among the group consisting of H, C₁₋₆ alkyl, C₂₋₆ alkenyl, C₂₋₆ alkynyl, C₄₋₈ alkadienyl, C₃₋₇ cycloalkyl, C₃₋₇ cycloheteroalkyl, aryl, and heteroaryl, said group being substituted with 0-3 R⁴, 0-3 R⁵ and 0-3 R⁹ or C₁₋₃ alkylene-NR⁴₂, C₁₋₃ alkylene-NR⁴C(O)R⁸, C₁₋₃ alkylene-NR⁴C(O)OR⁸, C₁₋₂ alkylene-O-NR⁴₂, C₁₋₂ alkylene-O-NR⁴C(O)R⁸, C₁₋₂ alkylene-O-NR⁴C(O)OR⁸ substituted with 0-3 R⁹,

- 35 FEP is a group selected among the group consisting of H, C₁₋₆ alkyl, C₂₋₆ alkenyl, C₂₋₆ alkynyl, C₄₋₈ alkadienyl, C₃₋₇ cycloalkyl, C₃₋₇ cycloheteroalkyl, aryl, and heteroaryl, said group being substituted with 0-3 R⁴, 0-3 R⁵ and 0-3 R⁹

or C₁-C₃ alkylene-NR⁴₂, C₁-C₃ alkylene-NR⁴C(O)R⁸, C₁-C₃ alkylene-NR⁴C(O)OR⁸, C₁-C₂ alkylene-O-NR⁴₂, C₁-C₂ alkylene-O-NR⁴C(O)R⁸, C₁-C₂ alkylene-O-NR⁴C(O)OR⁸ substituted with 0-3 R⁹,

where R⁴ is H or selected independently among the group consisting of

- 5 C₁-C₆ alkyl, C₂-C₆ alkenyl, C₂-C₆ alkynyl, C₃-C₇ cycloalkyl, C₃-C₇ cycloheteroalkyl, aryl, heteroaryl, said group being substituted with 0-3 R⁹ and

R⁵ is selected independently from -N₃, -CNO, -C(NOH)NH₂, -NHOH, -NHNHR⁶, -C(O)R⁶, -SnR⁶₃, -B(OR⁶)₂, -P(O)(OR⁶)₂ or the group consisting of C₂-C₆ alkenyl, C₂-C₆ alkynyl, C₄-C₈ alkadienyl said group being substituted with 0-2 R⁷,

- 10 where R⁶ is selected independently from H, C₁-C₆ alkyl, C₃-C₇ cycloalkyl, aryl or C₁-C₆ alkylene-aryl substituted with 0-5 halogen atoms selected from -F, -Cl, -Br, and -I; and R⁷ is independently selected from -NO₂, -COOR⁸, -COR⁸, -CN, -OSiR⁶₃, -OR⁶ and -NR⁶₂.

- R⁸ is H, C₁-C₆ alkyl, C₂-C₆ alkenyl, C₂-C₆ alkynyl, C₃-C₇ cycloalkyl, aryl or C₁-C₆ alkylene-aryl substituted with 0-3 substituents independently selected from -F, -Cl, -NO₂, -R³, -OR³, -SiR³₃
- 15

R⁹ is =O, -F, -Cl, -Br, -I, -CN, -NO₂, -OR⁶, -NR⁶₂, -NR⁶-C(O)R⁸, -NR⁶-C(O)OR⁸, -SR⁶, -S(O)R⁶, -S(O)₂R⁶, -COOR⁶, -C(O)NR⁶₂ and -S(O)₂NR⁶₂.

- 20 In a preferred embodiment Z is O or S, P is a valence bond, Q is CH, B is CH₂, and R¹, R², and R³ is H. The bond between the carbonyl group and Z is cleavable with aqueous I₂.

Partitioning conditions

- 25 The partition step may be referred to as a selection or a screen, as appropriate, and includes the screening of the library for encoded molecules having predetermined desirable characteristics. Predetermined desirable characteristics can include binding to a target, catalytically changing the target, chemically reacting with a target in a manner which alters/modifies the target or the functional activity of the target, and
- 30 covalently attaching to the target as in a suicide inhibitor.

- The target can be any compound of interest. E.g. the target can be a protein, peptide, carbohydrate, polysaccharide, glycoprotein, hormone, receptor, antigen, antibody, virus, substrate, metabolite, transition state analogue, cofactor, inhibitor, drug,
- 35 dye, nutrient, growth factor, cell, tissue, etc. without limitation. Particularly preferred

targets include, but are not limited to, angiotensin converting enzyme, renin, cyclooxygenase, 5-lipoxygenase, IL-10 converting enzyme, cytokine receptors, PDGF receptor, type II inosine monophosphate dehydrogenase, β -lactamases, integrin, and fungal cytochrome P-450. Targets can include, but are not limited to,

5 bradykinin, neutrophil elastase, the HIV proteins, including *tat*, *rev*, *gag*, *int*, RT, nucleocapsid etc., VEGF, bFGF, TGF β , KGF, PDGF, thrombin, theophylline, caffeine, substance P, IgE, sPLA2, red blood cells, glioblastomas, fibrin clots, PBMCs, hCG, lectins, selectins, cytokines, ICP4, complement proteins, etc.

10 Encoded molecules having predetermined desirable characteristics can be partitioned away from the rest of the library while still attached to the identifier nucleic acid sequence by various methods known to one of ordinary skill in the art. In one embodiment of the invention the desirable products are partitioned away from the entire library without chemical degradation of the attached nucleic acid identifier

15 such that the identifiers are amplifiable. The identifiers may then be amplified, either still attached to the desirable encoded molecule or after separation from the desirable encoded molecule.

In a preferred embodiment, the desirable encoded molecule acts on the target without any interaction between the nucleic acid attached to the desirable encoded

20 molecule and the target. In one embodiment, the bound complex-target aggregate can be partitioned from unbound complexes by a number of methods. The methods include nitrocellulose filter binding, column chromatography, filtration, affinity chromatography, centrifugation, and other well known methods.

25 Briefly, the library of complexes is subjected to the partitioning step, which may include contact between the library and a column onto which the target is immobilised. Identifier nucleic acids associated with undesirable encoded molecules, i.e. encoded molecules not bound to the target under the stringency conditions used, will pass

30 through the column. Additional undesirable encoded molecules (e.g. encoded molecules which cross-react with other targets) may be removed by counter-selection methods. Desirable complexes are bound to the column and can be eluted by changing the conditions of the column (e.g., salt, pH, surfactant, etc.) or the identifier.

- Additionally, encoded molecules which react with a target can be separated from those products that do not react with the target. In one example, a chemical compound which covalently attaches to the target (such as a suicide inhibitor) can be washed under very stringent conditions. The resulting complex can then be treated with proteinase, DNase or other suitable reagents to cleave a linker and liberate the nucleic acids which are associated with the desirable chemical compound. The liberated nucleic acids can be amplified.
- 10 In another example, the predetermined characteristic of the desirable product is the ability of the product to transfer a chemical group (such as acyl transfer) to the target and thereby inactivate the target. One could have a product library where all of the products have a thioester chemical group. Upon contact with the target, the desirable products will transfer the chemical group to the target concomitantly changing the desirable product from a thioester to a thiol. Therefore, a partitioning method
- 15 the desirable product from a thioester to a thiol. Therefore, a partitioning method which would identify products that are now thiols (rather than thioesters) will enable the selection of the desirable products and amplification of the nucleic acid associated therewith.
- 20 There are other partitioning and screening processes, which are compatible with this invention that are known to one of ordinary skill in the art. In one embodiment, the products can be fractionated by a number of common methods and then each fraction is then assayed for activity. The fractionization methods can include size, pH, hydrophobicity, etc.
- 25 Inherent in the present method is the selection of encoded molecules on the basis of a desired function; this can be extended to the selection of molecules with a desired function and specificity. Specificity can be required during the selection process by first extracting identifier nucleic acid sequences of chemical compounds which are capable of interacting with a non-desired "target" (negative selection, or counter-selection), followed by positive selection with the desired target. As an example, inhibitors of fungal cytochrome P-450 are known to cross-react to some extent with mammalian cytochrome P-450 (resulting in serious side effects). Highly specific inhibitors of the fungal cytochrome could be selected from a library by first removing
- 30 those products capable of interacting with the mammalian cytochrome, followed by
- 35

retention of the remaining products which are capable of interacting with the fungal cytochrome.

5

Brief Description of the Figures

Fig. 1 illustrates the overall process of building block evolution.

Fig. 2 shows the distribution of codon in different positions in an output from a selection.

10 Fig. 3 shows the difference between identifier driven and building block driven evolution.

Fig. 4 shows a method for reducing the library diversity through codon analysis.

Fig. 5 discloses two embodiments of using a Taqman probe (5' nuclease probe) in the measurement of the presence or absence of a certain codon.

15 Fig. 6 shows a standard curve referred to in example 4.

Fig. 7 shows a result of example 4.

Fig. 8 discloses a result of example 4.

Fig. 9 discloses a scheme relating to combined structural information and codon abundances in library design.

20 Fig. 10 discloses a relationship between codon analysis and structural information.

Fig. 11 shows the detection of single codons of identifiers.

Fig. 12 shows the detection of codon pairs of identifiers.

Fig. 13 shows the detection of codon pairs at specific codon positions.

25 Fig. 14 shows the detection of single codons of identifiers after the separation of the individual codons.

Fig. 15 discloses a method for selecting from a library, complexes capable of binding to a target molecule.

Fig. 16 discloses a method for enriching specific nucleic acid fragments and the utility of these fragments for the generation of a new library.

30 Fig. 17 discloses a method for reducing the diversity of a library of complexes.

Detailed description of the figures

Fig. 1A Shows the principle steps in BB evolution. An initial library of desired size is produced. This initial library is subjected to a selection process where encoded

35 molecules that associate with a target of interest are enriched. The encoding identi-

fier oligonucleotide is preferably amplified and the used in the codon analysis step. This step monitors the relative abundance of each codon in the selected library. The information obtained in this analysis is used to design a new enriched library, which contains the preferable chemical entities and their corresponding codons. This new library is then subjected to a new selection process to select for binders. This diversity reduction cycle can be repeated until the desirable result is obtained and the binders have been obtained.

Fig. 1B shows how the diversity of a library (n^4) is reduced by reducing the number of chemical entities (n) in the library. Thus, by removing chemical entities not involved in the encoded molecules partitioned, a reduction in library diversity can be obtained to allow the identification of binders.

The identifier oligonucleotide that encodes for the display molecule is composed of codons and associated with the encoded molecule, as shown in Fig. 2. These codons possess information about the chemical entities in the encoded molecule. Each of these codon positions can be analysed for the precise sequence, which will reflect which chemical entities that have been enriched for in the selection process. The relative amount can also be obtained by comparing the signal in the measuring procedure (e.g. QPCR and array analysis). Each codon position will have its own fingerprint on which chemical entities that the selected display molecules possess. These fingerprints in each position can subsequently be used to put together a new more focused library with a lower and more enriched diversity that can be subjected to another round of selection. This can then repeated until the preferable encoded molecules have been obtained.

Fig. 3 illustrates the main difference between identifier and chemical entity (CE) evolution. In both cases the initial selection starts on a library with certain diversity. After the first round of selection the encoding identifiers are amplified where the distribution is maintained. This distribution is then transferred to the next generation which is used in a new selection. Thus, the strongest binders that were enriched in the first round of selection will be present at a relatively higher concentration compared to the weaker binders and the background. In the CE-driven evolution the codon analysis is used to design a new library. In this example, the new library is constructed to contain all the chemical entities that were identified as a positive signal in

the analysis. In other words, all the chemical entities that were not detected through the codon analysis were excluded in the new library. The new library is designed to have an equal amount of each selected chemical entity, which will generate all the possible display molecules at the same concentration. This will allow all binders to compete at the same concentration and potentially retain a more diverse set of binders in each round of selection. This is especially important for small molecules here not only the affinity is of interest.

Figure 4. This illustrates the process where the diversity is reduced through the codon analysis. An initial library of 10^{10} (e.g. 317*317*317*317) library members is subjected to a selection. The enrich identifier oligonucleotides are amplified and used in the codon analysis. The codon analysis result is used to design a new 10^7 (e.g. 57*57*57*57) library where the enriched chemical entities are included. This new library is the again subjected to a selection process. The identifier oligonucleotides are amplified and used for codon analysis. This new codon analysis results is again used to design a new 10^4 (e.g. 10*10*10*10) library where the enriched chemical entities are included. Finally a last selection step is performed in this reduced diversity library to identify the binders.

A preferred embodiment of the invention utilizing a universal Taqman probe is shown in Fig.5. Four codons are shown (P1 through P4; bold pattern) along with flanking regions (light pattern). A universal Taqman probe anneals to a region adjacent to the codon region, but within the amplicon defined by the universal PCR primers Pr.1 and Pr. 2. These primers could be the same as used for amplification of the identifier oligonucleotides encoding binders after an enrichment process on a specific target. However, are minimal length identifiers preferred during the encoding process, the region involved in Taqman probe annealing could be appended to the library identifier oligonucleotides by e.g. overlap PCR, ligation, or by employing a long downstream PCR primer containing the necessary sequences. The added length corresponding to the region necessary for annealing of the Taqman probe would be form 20 to 40 nts depending on the type of TaqMan probe and T_A of the PCR primers. The Q-PCR reactions are preferably performed in a 96- or 384-well format on a real-time PCR thermocycling machine.

- Panel A shows the detection of abundance of a specific codon sequence in position one. Similar primers are prepared for all codon sequences. For each codon sequence utilized to encode a specific BB in the library a Q-PCR reaction is performed with a primer oligonucleotide complementary to the codon sequence in question. A downstream universal reverse primer Pr. 2 is provided after the Taqman probe to provide for an exponential amplification of the PCR amplicon. The setup is most suited for cases where the codon constitutes a length corresponding to a length suitable for a PCR primer.
- Panel B shows the detection of abundance of a specific codon sequence in a specific codon position using a primer, which is complementing a codon and a framing sequence. Similar primers are used for all the codons and framing sequences. For each codon sequence utilized to encode a specific BB at a specific codon position in the library a Q-PCR reaction is performed with an oligo complementary to the codon sequence in question as well as a short region up- or downstream of the codon region which ensures extension of the primer in a PCR reaction only when annealed to the codon sequence in that specific codon position. The number of specific primers and Q-PCR reactions needed to cover all codon sequences in all possible codon positions equals the number of codon sequences times the number of codon positions. Thus, monitoring the abundance of 96 different codon sequences in 4 different positions can be performed in a single run on four 96 wells micro titre plates (as shown in Panel B) or a single 384 well plate on a suitable instrument. This architecture allows for the decoding of a $8,5 \times 10^7$ library of different encoded molecules.
- Quantification is performed relative to the amount of full-length PCR product obtained in a parallel control reaction on the same input material performed with the two external PCR primers Pr.1 + Pr. 2. Theoretically, a similar rate of accumulation of this control amplicon compared to the accumulation of a product utilizing a single codon + sequence specific primer would indicate a 100% dominance of this particular sequence in the position in question.
- Although the setups shown in Panel A and B employ a Taqman probe strategy, other detection systems (SYBR green, Molecular Beacons etc.) could be utilized. In theory, multiplex reactions employing up to 4 different fluorophores in the same reaction could increase throughput correspondingly.

An example of how a deconvolution process of a library of encoded molecules occurs is described in the following. Imagine that at the end of a selection scheme a pool of 3 ligand families (and the corresponding coding identifiers) are dominating the population and present at approx. the same concentration. Three different chemical entities are present in the first position of the encoded compounds, and each of these chemical entities are present in combination with one unique chemical entity out of 3 different chemical entities in position P2. Only one chemical entity in position 3 gives rise to active binders, whereas any of a 20% subset of chemical entities (e.g. determined by charge, size or other characteristics) is present in position 4. The outcome of the initial codon profile analysis would be: 3 codon sequences are equally dominating in position P1, 3 other codon sequences in position P2, 1 unique codon sequence is dominant in P3 whereas somewhat similarly increased levels of 20% of the codon sequences (background levels of the remaining 80% sequences) are seen in P4. In such cases it could be relevant to use an iterative Q-PCR ("IQ-PCR") strategy to perform a further deconvolution of a library after selection. Again with reference to the example above, by taking the PCR products from the 3 individual wells that contained primers giving the high yields in position P1, diluting the product appropriately and performing a second round of Q-PCR on each of these identifier oligonucleotides separately, it would be possible to deduce which codon sequence(s) is preferred in P2 when a given codon sequence is present in P1.

Fig. 9. This figure illustrates the possibility to combine structural information about the chemical entities and the relative abundance when designing a new more focused library. The structural information about the chemical entities can be used at least in two ways. First the similarities between the chemical entities in each position can be used to choose chemical entities to a new library. Secondly, the combination of the selected chemical entities can be analyzed to investigate possible pattern that generate potential ligands. This is especially useful if the binding site or the structure of a known ligand is known. Any type of structural analysis tool can be used that generate information about the structure of separate chemical entities or combination of chemical entities (the potential binders). By combining these three analysis approaches a more focused library can be generated that potentially will contain more specific binders compare to background binders. This new focused library can

be used in another round of selection to reduce the diversity. This procedure can be repeated until the desired binders have been identified.

Figure 10. This figure shows how the combination of codon analysis and structural information can generate valuable information. This invention allows the performance of structure activity relationship analysis (SAR) where the relative abundance in the codon analysis will represent the activity parameter (e.g. IC_{50} values) in the SAR measurements. Pharmacophore models can be generated, focused libraries can be designed, certain follow up chemistry can be used and information in the hit to lead process can be used.

Fig. 11 shows an array detection system in which a single codon is detected. Initially a library of selected complexes (29), i.e. complexes comprised of the initial library, which display a certain property, is provided as disclosed above. The initial library of complexes is prepared from e.g. 100 codons and identifiers having 4 codons in sequence, which theoretical gives a library of 10^8 complexes. The selected complexes are subjected to amplification to amplify the identifiers of the selected complexes and the amplification products are added to an array (30). The array (30) comprises probes (32) complementary to each of the codons of the identifiers (31). At hybridisation conditions the PCR products of the identifiers are annealed to the cognate probes of the array and in a suitable scanner the spatial position of the annealed probes are detected to elucidate the codons (33) of the identifier. The quantity of each codon may be measured to find codons abundant in more than one identifier and/or codons leading to encoded molecules with high affinity. The information may be used for decoding of the encoded molecule of the complexes displaying the desired property or the information may be used for selection of building blocks, which is to be added in a next round of library formation.

Fig. 12 discloses an array detection system for establishing codons pairs, i.e. codons in the vicinity of each other. Initially (as shown in this example) a library of complexes is prepared from 100 different codons deposited on an identifier in a sequence of four, making the total amount of combinations possible 10^8 . The initial library is subjected to a condition in order to select a sub-library (29) displaying a desired property. The identifiers of the sub-library are amplified by a PCR reaction and the reaction product is added under hybridisation conditions to an array (34).

The array is designed with probes (35) capable of detecting two codons at a time. To cover all possible combinations of a library based on 100 different codons 10^4 probes are needed, which is practically feasible with the current technology.

5 The detection of the codons may be conducted quantitatively, i.e. the relative abundance of each of the codon pairs may be determined. The detection on the array may be used to reconstruct the selected identifiers (36) as three overlapping codon pair detections depict the entire identifier. In the event the same codon pair appears on more than one identifier, the information on the relative abundance of each codon pair maybe used to decipher the sequence of codons of the selected identifiers as it can be assumed that each codon pair of the same identifier appears in the same amounts in the PCR products added to the array.

Fig. 13 discloses an array for detecting codon pairs at specific codon positions. Initially, a library of complexes comprising identifiers with framing sequences is provided. The framing sequence is specific for each position of the codons on the identifier. Four times more probes on the microarray is needed per each codon if the position of the codons also should be detected in the analysis which is practically feasible with current technology. The position is detected due to the framing sequences next to each codon. The initial library is subjected to a selection process to isolate complexes (37) having a desired property. The selected complexes are amplified by a PCR reaction and the reaction products are added to an array (38). The array comprises probes capable of detecting codon pairs as wells as the framing sequences (40) between the codons. The framing sequence determines the position of the codon in the reaction history, i.e. it is possible to deduct which chemical entity that reacted at which point in time of the synthesis history of the encoded molecule, thus making it possible to reconstruct the structure of the encoded molecule.

The detection of the codon pairs may be conducted quantitatively, i.e. the relative abundance of each of the codon pairs may be determined. The detection on the array may be used to reconstruct the selected identifiers (41) as three overlapping codon pair detections depict the entire identifier. In the event the same codon pair appears on more than one identifier, the information on the relative abundance of each codon pair maybe used to decipher the sequence of codons of the selected identifiers as it can be assumed that each codon pair of the same identifier appears in the same amounts in the PCR products added to the array.

Fig. 14 shows an array detection system in which a single codon is detected. Initially a library of selected complexes (42), i.e. complexes comprised of the initial library which display a certain property, is provided as disclosed above. The initial library of complexes is prepared from e.g. 100 codons and identifiers having 4 codons in sequence, which theoretical gives a library of 10^8 complexes. The selected complexes are subjected to amplification to amplify the identifiers of the selected complexes and the amplification products are treated with suitable reagents to cut between the individual codons (43). The individual codon is then applied to the array. The array (44) comprises probes (45) complementary to each of the codons of the identifiers (46). At hybridisation conditions the PCR products of the identifiers are annealed to the cognate probes of the array and in a suitable scanner the spatial position of the annealed probes are detected to elucidate the codons (47) of the identifier. The quantity of each codon may be measured to find codons abundant in more than one identifier and/or codons leading to encoded molecules with high affinity. The information may be used for decoding of the encoded molecule of the complexes displaying the desired property or the information may be used for selection of building blocks, which is to be added in a next round of library formation.

Fig. 15 discloses a method for selection of a suitable complex in several steps. In a first step the library of complexes 1 is provided. Each member of the library comprises an encoded molecule 2 composed of four chemical entities which is attached to an identifier oligonucleotide 3, which comprises four codons. The initial library shown comprises three complexes. In a second step the library of complexes is incubated with immobilized target molecules 4. The encoded molecule having an affinity towards the target molecule is bound to the immobilized target whereas encoded molecules not having affinity towards the target under the conditions used remains in the liquid media. The complexes remaining in the liquid media are discarded by a washing process, while the bound complexes remain attached to the immobilized target molecules. The washing process is usually conducted using mild stringency conditions in the initial rounds of selection. In later stage selections the working stringency conditions are usually increased to allow only high affinity binders to remain attached to the target. Subsequent to the washing step the complexes having affinity towards the target molecule are recovered. The recovery process usually requires high stringency conditions to detach the encoded molecule from the

immobilized the target. The selected sub-library resulting from the elution is subjected to an amplification process. The amplification of the identifier nucleic acid sequence of the selected complexes is usually performed using the PCR method. Preferably, a modification of the PCR method is followed such that a biotin molecule
5 is attached to one of the primers to obtain a handle for subsequent immobilization. The result of the amplification step is multiple copies of the identifier nucleic acid sequences, which codes for the encoded molecules which have survived the selection step.

- 10 Fig. 16 discloses an enrichment process of building blocks. The building blocks can be used for generation of a new library. Initially, identifier nucleic acid sequences are immobilized on solid support. In one aspect of the invention the identifier nucleic acid sequences are the product of the selection procedure described in Fig. 1. Each codon of the identifier nucleic acid sequence is identified with an uppercase letter,
15 i.e. A, B, C, or D. The immobilized identifier acid sequences are contacted with the pool of building blocks under hybridisation conditions. Each of the building blocks are illustrated with an sequence complementary to a codon which may or may not be present on the identifier nucleic acid sequence. The complementary sequences are indicated with a apostrophe, e.g. A', B', etc. The transferable chemical entity of a
20 building block is illustrated with a lowercase letter. The conditions providing for hybridisation of the complementing sequences of the pool of building blocks to the immobilised identifier nucleic acid sequence are preferably such that cognate nucleic acid sequences are hybridised to each other while sequences not recognizing any immobilized sequence remain in aqueous media. The immobilized sequences of
25 the identifier nucleic acid sequences are thus used as bait in catching building blocks with complementing sequences. Following the incubation step, non-binding building blocks are removed by washing, whereby the part of the pool of building blocks not being able to find a complementing sequence is discarded. The building blocks attached to the immobilized nucleic acid sequences are detached using de-
30 hybridisation conditions. The diminished pool of building blocks may be used in a subsequent round for preparing a new library of complexes, in which the encoded molecule comprises a reaction product comprising additions from chemical entities attached to the enriched building blocks. Because the order of building blocks which have participated in the formation of the encoded molecules successful in the selection procedure, is not preserved by the method for enriching building blocks a
35

scrambling of the encoded molecules may be obtained in some of the methods described herein for obtaining a library of complexes. In some applications of the library it will be an advantage to have a scrambling of the building blocks because and increased diversity is obtained.

5

Fig. 17 discloses a method for reducing the diversity of the library of complexes resulting from the method described in Fig. 16. In some of the applications of the library the diversity induced by scrambling of the building blocks are not desired. In a first step the sequences complementary to the identifier acid sequences used in Fig. 10 16 are provided and immobilized on a suitable solid support. In one aspect of the invention the complementary sequence is obtained from the PCR product resulting from the method according to Fig. 15. Alternatively, the complementing sequence may be obtained by extending the identifier nucleic acid sequence using a suitable primer, optionally attached to a handle such as a biotin or dinitrophenol. In a second 15 step the immobilized complementary sequence is incubated with the scrambled library under conditions, which provide for hybridisation between the complementary sequence and members of the library having affinity towards this sequence. Members of the library not having affinity to the complementary sequences remains in the media and is discarded, while members of the library being able to hybridise to the 20 immobilized nucleic acid sequences is recovered. Occasionally, nucleic acids not perfectly matching with the complementary sequence immobilized on the solid support are caught. In one aspect of the invention the hybridisation products, prior to the recovery step, are treated with an enzyme capable of recognizing mismatching nucleotides and cleaving the double stranded helix in which they are situated. An example of an enzyme with this ability is T4 endonuclease VII. After the treatment with 25 the enzyme, complexes displaying a hybridisation toward the immobilized sequence are eluted under dehybridisation conditions. Nucleotide sequences remaining from the cleavage by the enzyme will also be present in the new library, however, these sequences will not have any effect of a subsequent selection because no molecule 30 is attached thereto.

Examples

Example 1. Enrichment of nucleic acid fragments

A codon was included in the oligonucleotide sequence shown below. The codon is underlined and the boldface sequences represent the "framing" regions next to each codon. These framing regions can be used for specifying the position of each codon.

Biotin-AATTCCGGAACATACTAGTCAACATGA-3' (SEQ ID NO:1)

This identifier oligonucleotide was immobilized on streptavidin beads using standard protocols, i.e. 600 pmol identifier oligonucleotide with 5'- dT biotin in 50 µl 100 mM Mes pH 6.0 was mixed with 50 µl SA-magnetic beads (Roche). The mixture was washed 2-3 times with 100 mM MES pH 6.0 to remove non-bound identifier oligonucleotides. To reduce background binding, the oligos and beads were incubated at RT for 10 min on shaker, then incubated on ice for 10 min while rotating the tube. Finally, the sample was washed with 100 mM MES 4 times in 800 µl at 60°C.

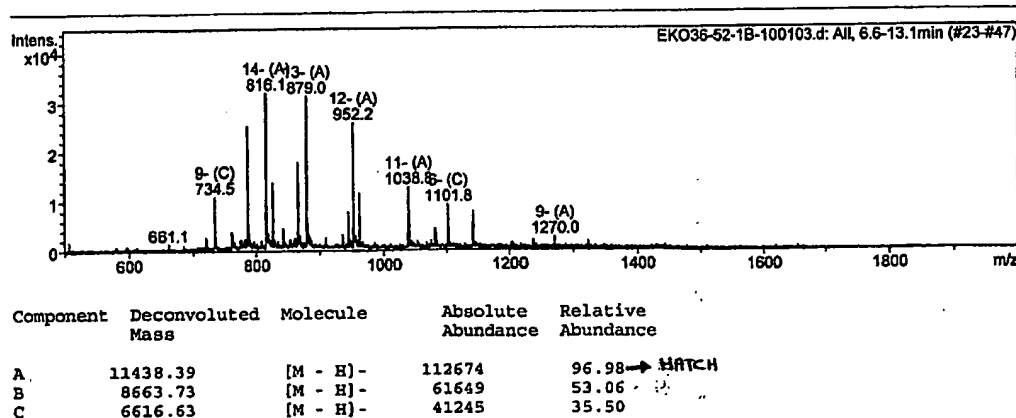
In the case where a PCR product is immobilized, the complementing (non-sense) strand is removed using 10 mM NaOH. This will generate single-stranded DNA with the selected codons. The same procedure described in this example can be used for a collection of different identifier nucleic acid molecules that contain one or more codons. The codons in the identifier nucleic acid molecules can be the same or different determined from the enrichment performed on the initial library.

The immobilized identifier nucleic acid molecule was mixed with the pool of nucleic acid fragments shown below. This pool of fragments illustrates an original pool that was used for generating an initial library of complexes. Each fragment may possess in the 3'-end a specific chemical entity that is encoded by the codon sequence. These nucleic acid fragments contain a specific sequence in the codon region (underlined) while the framing region shown in boldface is identical among the fragments. Thus, the pool of fragments represents different codons in the same position of the identifier nucleic acid.

1. CGT GTG ATC GAA CTC GTG TG GTATGATCAGTTGTACT-5'
(SEQ ID NO:2)

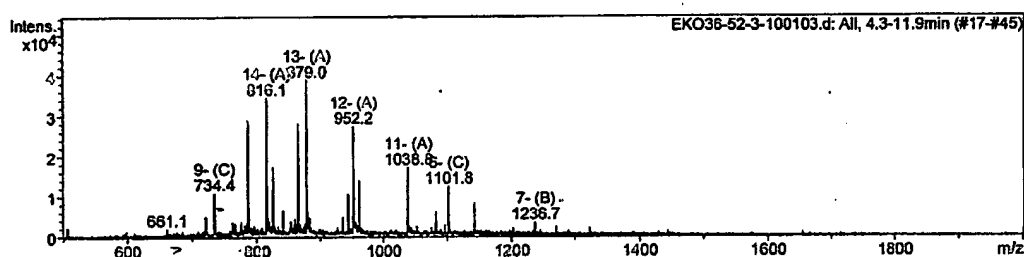
2. CGT GTG ATC GAA CTC GTG TG GTATCTAGTCGGTTACT-5'
(SEQ ID NO:3)
3. CGT GTG ATC GAA CTC GTG TG GTATTCGAGTGTTTACT-5'
(SEQ ID NO:4)
- 5 4. CGT GTG ATC GAA CTC GTG TG GTATAGCTCATGGTTACT-5'
(SEQ ID NO:5)

The nucleic acid fragments are mixed with the immobilized identifier nucleic acid using 600 pmol of each nucleic acid fragment mixed with the immobilized identifier nucleic acid molecules (100 mM MES pH 6.0, 150 mM NaCl)). The mixture was incubated at 25°C for 30 minutes in a shaker. The non-hybridized fragments were removed by 4 times washing in 800 μ l 100 mM MES, 150 mM NaCl. This step should separate the complementing fragments (bound) encoding for the select chemical entities from the non-complementing fragments (non-bound) encoding for chemical entities that were not effective in the preceding selection process. The annealed fragments are eluted from the immobilized identifier nucleic acid molecules by re-suspending the beads in 25 μ l 60°C H₂O and incubating for 2 min at 60°C. The enriched fragments were purified on a micro-spin gel filtration column (BiRad). The eluted fragments were prepared for mass spectroscopy (MS) analysis by mixing in half volume of ion exchanger resin and incubating minimum 2 h at 25°C on a shaker. After incubation the resin was removed by centrifugation and 15 μ l of the supernatant was mixed with 7 μ l of water, 2 μ l of piperidine and imidazole (each 625 mM) and 24 μ l acetonitrile. The sample was analysed using a Mass Spectroscopy instrument (Bruker Daltonics, Esquire 3000plus). The result for the MS analysis is shown below.



The mass from the correct complementary fragment (number 1) is obtained in the MS analysis (11438.39, expected 11439 Da) No masses for the other fragments (number 2-4) could not be found in the MS spectra (expected masses; 11415, 11430, 11424 Da). This result shows that the right fragment is strongly enriched and other fragments with the wrong codon sequences are removed. The enrichment is possible even when the "spacing" region (boldface) is identical in each fragment.

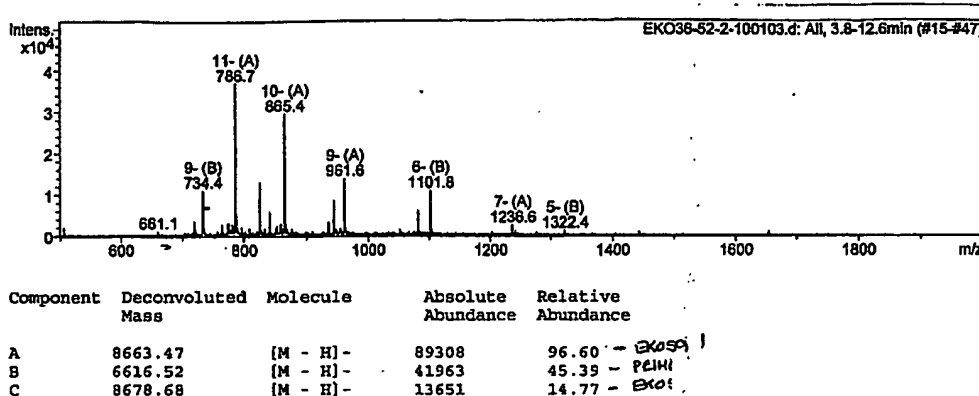
Two control experiments were also performed to validate the enrichment protocol. In the first experiment, the fragment with the correct codon sequence (number 1) was mixed with the immobilized identifier molecule as described above. The sample was washed and eluted also as described above and prepared for MS analysis. The result from the MS analysis is shown below.



Component	Deconvoluted Mass	Molecule	Absolute Abundance	Relative Abundance
A	11438.39	[M - H] -	127110	96.21
B	8663.61	[M - H] -	82800	62.67
C	6616.52	[M - H] -	51480	38.97

The result indicates that the fragment with the correct sequence (number 1) anneals to the immobilized identifier molecules and is eluted under the conditions used in this example. The expected mass (11439) correlate well with the experimental mass, 11438.39 Da.

In the other control experiment, a fragment with a wrong codon sequence (number 3) was allow to bind to the immobilized identifier molecule as described above. Again, the eluted sample was prepared and analysed with MS. The result is shown below.



In this experiment, no mass was found that corresponded to the expected mass (11430) of the tested building block (number 1). Again, this shows that fragments with a anticodon sequence different from the enriched codons in the identifier nucleic acid molecules are not captured using this approach.

The enriched fragments obtained using this strategy may then be used to generate a new library of encoded molecules. This new library will contain encoded molecules composed of the enriched chemical entities. Thus, the library size have been reduced due to the removal of chemical entities not involved in binding encoded molecules, and enriched in chemical entities that are highly represented in the encoded molecules which binds to the target molecule.

Example 1 shows the possibility of enriching for specific building block molecules, i.e. nucleic acid fragments associated with transferable chemical entities. The same procedure can be used for a larger pool of building block than four as used herein. The codon design will determine the maximum number of building blocks that can be used. The sequence in the codon region should be large enough to allow discrimination in the annealing step. Various conditions can be used to increase the stringency in the annealing step. Parameters such as temperature, salt, pH, formamide concentration, time and other conditions could be used.

Example 2 (model): Multiple codon selection in a library.

This example describes the enrichment of building blocks using an identifier nucleic acid (identifier) molecule with multiple codons. These codons encode for a displayed molecule (DM) that are attached to the identifier molecule before the selection is

performed. The library size is determined both by the number of different chemical entities and the total number of chemical entities. The identifier molecule shown below contains three codons. The codons, which codes for the displayed molecule are indicated with underlines and the region separating (framing region) the codons are indicated with underlines and the region separating (framing region) the codons in boldface. The size of the codons can be varied dependent in the diversity need in the library and the optimal setup for chemical entity enrichment. The framing region can also be varied dependent on the discrimination needed to distinguish the precise position of a codon in the identifier molecule. The framing region will also be important for the generation of the library. This can be understood when the encoding is accomplished by extension of the encoding region as disclosed in DK PA 2002 01955 and US 60/434,425, incorporated herein by reference. There need to be a perfect match in the 3'-end in order to get efficient extension with a polymerase or a ligase. The size of this spacing/framing region should be long enough to form a complementing region to allow extension with a polymerase or ligase. Preferably, the spacing region should be between 3 and 6 nucleotides. The codon region together with the spacing region will also be useful when codons are to be identified using a micro array setup. The identifier molecule with the right codon sequences will hybridize to the array and be detected.

The sequence below represents an enriched identifier molecule attached to the displayed molecule (DM). This identifier molecule has been enriched due to the fact that the DM binds to the target molecule in the selection process. In practice, more than one enriched identifier molecules will be obtained when using a library of displayed molecules attached to its identifier sequence.

DM-

GCACACTAGCTTGAGCACACTGACACATGGAGATCACATGCTTCGACAATGCAGGACTCCCGCAGCTTTACGATCCCGCAGGTAACCGT

This identifier molecule is amplified with two primers (below) using a standard PCR reaction. For example, 500 nM of each primer, 2,5 units Taq polymerase, 0.2 mM of each NTP, in a PCR buffer (50 mM KCl, 10 mM Tris-Cl, 3 mM DTT, 1.5 mM MgCl₂, 0.1 mg/ml BSA). Run 25 cycles (94°C melt for 30 seconds, 55°C anneal for 45 seconds, 72°C extension for 60 seconds).

B-GCACACTAGCTTGAGCACACTGACA-3'

CGAAATGCTAGGGCGTCCATTGGCA-5'

This will amplify the identifier molecule from the selection process and add a biotin in the 5'-end of one of the strand (below). This amplified product is then immobilized on a solid support, streptavidin beads for example. This can be performed identical as describe in example 1.

5

When the identifier molecules have been immobilized and the excess has been removed by a washing step (as describe in example 1), the complementing non-sense stand is removed by incubating in 10 NaOH for about 2 min and washed with 100 mM Mes buffer, pH 6.0. This procedure will generate the strand shown below where

10 the codon regions are exposed to allow hybridization with the complementing sequences.

B-

GCACACTAGCTTGAGCACACTGACACATGGAGATCACATGCTTCGACAATGCAGGACTCCCGCAGCTTTACGATCCCGCAGGTAACC

15 GT

The next step is to protect the complementing sequences outside the codons to prevent the binding of the building block to these sequences. This can be performed by adding "blocking" oligonucleotides that has a complementing sequence. This is

20 shown below.

B-

GCACACTAGCTTGAGCACACTGACACATGGAGATCACATGCTTCGACAATGCAGGACTCCCGCAGCTTTACGATCCCGCAGGTAACC

GT

25 CGTGTGATCGAACTCGTGTGACTGT

CGAAATGCTAGGGCGTCCATT-

GGCA

Next, the pool of different building blocks is added and is allowed annealing to the codon region in the identifier region. The position of annealing is determined by the

30 spacing region shown in boldface. The stringency is adjusted to only allow hybridization of the correct building block in the right position. This can be accomplished by mixing the right component together using various conditions. The condition can for example include the presence of salt, formamide and various buffers adjusted to suitable pH and temperature. Below is the correct building block that will anneal to

35 the enriched identifier molecules. These building blocks is annealed and eluted as described in example 1.

CE-CGTGTGATCGAACTCGTGTGACTGTGTACCTCTAGTGTAC

The next pool of building blocks is blocked with an oligonucleotide that also protects the first codon. This is necessary to prevent binding of the building blocks in that codon.

5 B--
 GCACACTAGCTTGAGCACACTGACACATGGAGATCACATGCTTCGACAAATGCAAGGACTCCCGCAGCTTTACGATCCCGCAGGTAACC
 GT
 CGTGTGATCGAACTCGTGTGACTGTGTATTTTTTTTTT CGAAATGCTAGGGCGTCCATT-
 GGCA

10 Again, the library of building blocks is added to enrich for the selected codons. Below is the building block with the correct sequence. These building blocks is annealed and eluted as described in example 1.

15 CE--CGTGTGATCGAACTCGTGTGACTGTGTATTTTTTTTTTTACGAAGCTGTTACG

Finally, the identifier molecule is protected with a blocking oligo that expose only the last codon.

20 B--
 GCACACTAGCTTGAGCACACTGACACATGGAGATCACATGCTTCGACAAATGCAAGGACTCCCGCAGCTTTACGATCCCGCAGGTAACC
 GT
 CGTGTGATCGAACTCGTGTGACTGTGTATTTTTTTTTTTACIIIIIIIIII CGAAATGCTAGGGCGTCCATT-
 GGCA

25 A new pool of building blocks is added and allowed hybridizing to the identifier molecule. These building blocks is annealed and eluted as described in example 1.

CE--CGTGTGATCGAACTCGTGTGACTGTGTATTTTTTTTTTTACIIIIIIIIIIACGTCCTGAGGGCGT

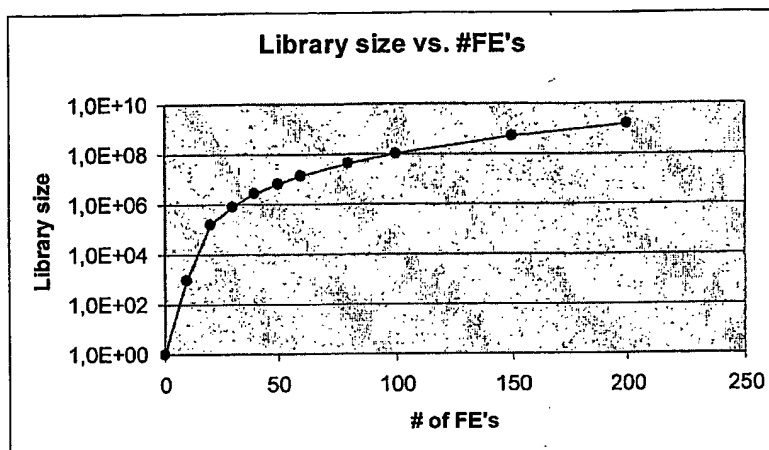
30 The enrichment of each library of building blocks are performed in separate tubes in order to keep the libraries of building block separated. The enrichment is performed with building blocks loaded with chemical entities (CE).

35 Example 3 – Template versus chemical entity evolution

The graph below illustrates the relationship between the number of chemical entities and the library size. The example below is calculated on that the final encoded molecules contains four chemical entities that is individually encoded by the corresponding building block (n^4 , where n is the number of building blocks). The graph shows that the diversity decreases dramatically with the reduction of the total number of building blocks. If the number of different building can be reduced to about 20-

40

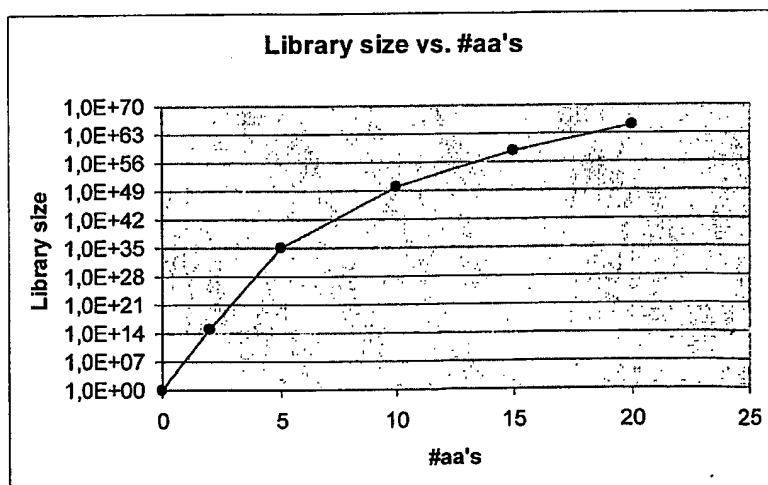
30 (library size of $16 \cdot 10^3$ and $81 \cdot 10^4$, respectively) in the selection process, then the library size for the final round of selection is low enough for identification of the binding molecules.



5

When the same analysis is performed on a protein another situation is obtained. The example shown below is on a very small protein (50 amino acids in length). The diversity is enormous when all amino acids are included in the library. The size of the library is also decreasing with the total number of amino acids, but not to the same extent as show above for a small molecule. Even when the different amino acids are reduced to 2, the library size is huge ($1.2 \cdot 10^{15}$). This shows that amino acid enrichment is impossible on protein. This is even more pronounced for mid-size protein which contains about 300 amino acids.

15



Example 4 – Codon analysis

This example illustrates one possibility to perform codon analysis on a whole population of different identifier oligonucleotides. The analysis can also be performed using array where the probe oligonucleotides (complementary to the codons) are immobilized in discreet areas and the signal is monitored dependent on the amount of identifiers oligonucleotides are hybridised in each specific area. Codon analysis can also be performed using standard sequencing using a polymerase extension step.

10

In fig. 5, Four codons are shown (P1 through P4; bold pattern) along with flanking regions (light pattern). A universal Taqman probe anneals to a region adjacent to the codon region, but within the amplicon defined by the universal PCR primers Pr.1 and Pr. 2. These primers could be the same as used for amplification of the identifier oligonucleotides encoding binders after an enrichment process on a specific target. However, are minimal length identifiers preferred during the encoding process, the region involved in Taqman probe annealing could be appended to the library identifier oligonucleotides by e.g. overlap PCR, ligation, or by employing a long downstream PCR primer containing the necessary sequences. The added length corresponding to the region necessary for annealing of the Taqman probe would be from 20 to 40 nts depending on the type of Taqman probe and T_A of the PCR primers. The Q-PCR reactions are preferably performed in a 96- or 384-well format on a real-time PCR thermocycling machine.

25 Fig. 5, panel A, shows the detection of abundance of a specific codon sequence in position one. Similar primers are prepared for all codon sequences. For each codon sequence utilized to encode a specific BB in the library a Q-PCR reaction is performed with a primer oligonucleotide complementary to the codon sequence in question. A downstream universal reverse primer Pr. 2 is provided after the Taqman probe to provide for an exponential amplification of the PCR amplicon. The setup is most suited for cases where the codon constitutes a length corresponding to a length suitable for a PCR primer.

35 Fig. 5, panel B shows the detection of abundance of a specific codon sequence in a specific codon position using a primer which is complementing a codon and a fram-

ing sequence. Similar primers are used for all the codons and framing sequences. For each codon sequence utilized to encode a specific BB at a specific codon position in the library a Q-PCR reaction is performed with an oligo complementary to the codon sequence in question as well as a short region up- or downstream of the codon region which ensures extension of the primer in a PCR reaction only when
5 annealed to the codon sequence in that specific codon position. The number of specific primers and Q-PCR reactions needed to cover all codon sequences in all possible codon positions equals the number of codon sequences times the number of codon positions. Thus, monitoring the abundance of 96 different codon sequences
10 in 4 different positions can be performed in a single run on four 96 wells micro titre plates (as shown in Fig. 5, panel B) or a single 384 well plate on a suitable instrument. This architecture allows for the decoding of a $8,5 \times 10^7$ library of different encoded molecules.

15 Quantification is performed relative to the amount of full-length PCR product obtained in a parallel control reaction on the same input material performed with the two external PCR primers Pr.1 + Pr. 2. Theoretically, a similar rate of accumulation of this control amplicon compared to the accumulation of a product utilizing a single codon + sequence specific primer would indicate a 100% dominance of this particular sequence in the position in question.
20

Although the setups shown in Fig. 5, panel A and B employ a Taqman probe strategy, other detection systems (SYBR green, Molecular Beacons etc.) could be utilized. In theory, multiplex reactions employing up to 4 different fluorophores in the same
25 reaction could increase throughput correspondingly.

An example of how a deconvolution process of a library of encoded molecules occurs is described in the following. Imagine that at the end of a selection scheme a pool of 3 ligand families (and the corresponding coding identifiers) are dominating
30 the population and present at approx. the same concentration. Three different chemical entities are present in the first position of the encoded compounds, and each of these chemical entities are present in combination with one unique chemical entity out of 3 different chemical entities in position P2. Only one chemical entity in position 3 gives rise to active binders, whereas any of a 20% subset of chemical
35 entities (e.g. determined by charge, size or other characteristics) are present in posi-